DOCUMENT RESUME

ED 147 923 EA 010 113

AUTHOR Wright, Edgar N.; And Others
TITLE Effects of Class Size in the Junior Grades. A Study.
INSTITUTION Ontario Dept. of Education, Toronto.
PUB DATE 77
NOTE 152p.; Occasional pages contain very small print
AVAILABLE FROM The Ontario Institute for Studies in Education, Publications Sales, 252 Bloor Street West, Toronto, Ontario, Canada M5S 1V6 ($2.00 bound; $1.50 microfiche)

EDRS PRICE MF-$0.83 HC-$8.69 Plus Postage.
DESCRIPTORS *Academic Achievement; *Class Size; Elementary Education; *Graphs; Parent Attitudes; Statistical Analysis; Student Attitudes; *Student Teacher Ratio; *Student Teacher Relationship; *Tables (Data); Teacher Attitudes
IDENTIFIERS *Ontario (Toronto)

ABSTRACT

This study investigated the differences between different sized classes in grades four and five of Toronto schools. Teachers and students from 11 schools were assigned to 34 classes of either 16, 23, 30, or 37 students (class size could vary by 2). All students were in grade four during the study's first year and in grade five during the second year. Questionnaires and attitude scales were used to measure teachers' expectations and attitudes, students opinions and attitudes toward their classes, and parents' opinions. In addition, student achievement in the areas of reading comprehension, vocabulary, mathematics problem-solving, mathematics concepts, art, and composition was measured, and students' academic self-concepts were assessed. Results of the study showed that varying class size produced few changes in classroom functioning, and most differences between class sizes were relatively small. Students were addressed as individuals significantly more frequently as class size was reduced below 30. Other differences included increased mathematics concepts scores in classes of 16, increased verbal participation in classes of 16 or 23, more noise in classes of 30 or 37, increased scores on "indicators of quality" in classes of 16 or 23, more frequent reading instruction in classes of 16, and increased complaints about lack of art instruction in classes of 37. (Author/JG)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED147923

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EFFECTS OF CLASS SIZE IN THE JUNIOR GRADES

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Ontario Ministry of Education

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

A Study

EDGAR N. WRIGHT, Principal Investigator
STANLEY M. SHAPSON
GARY EASON
JOHN FITZGERALD

This research project was funded under contract by the Ministry of Education, Ontario

EA 010 113

© The Minister of Education, Ontario 1977
Queen's Park
Toronto, Ontario

Additional copies may be ordered from:

The Ontario Institute for Studies in Education
Publications Sales
252 Bloor Street West
Toronto, Ontario
M5S 1V6

or from:

The Ontario Government Bookstore
880 Bay Street
Toronto, Ontario
M7A 1L2

MINISTRY OF EDUCATION, ONTARIO CATALOGUING IN PUBLICATION DATA

Wright, Edgar N
Effects of class size in the junior grades : a study / [by] Edgar N. Wright, principal investigator . . . [et al.]
Toronto : Minister of Education, Ontario, 1977.

"This research project was funded under contract by the Ministry of Education, Ontario."
130 p.
1. Class size I. Wright, Edgar N. II. Toronto Board of Education. III. Ontario. Ministry of Education. IV. Title.
ON00691 371.251

1 2 3 4 5 XX 18 08 97 87 77

Abstract

Previous research on the effects of class size is difficult to interpret because researchers have often failed to control important variables and have not agreed on the definitions of "large" and "small" classes. This study attempted to investigate experimentally the differences between four class sizes which ranged from the size of many special education classes to the largest sized regular classroom found in Toronto schools.

In the first year of the study, teachers and Grade 4 students from 11 schools in Metropolitan Toronto were randomly assigned to 34 classes of 16, 23, 30 or 37 (class sizes could vary by 2); student assignment was stratified by sex and by academic performance. In the second year the same teachers and students were similarly assigned to Grade 5 classes, with the constraints that students should not be in a class of 16 or of 37 for both years of the study, and that teachers should not have a class of the same size both years. Students came from all socio-economic categories, but the lower categories were over-represented in comparison to the Toronto school population.

The teachers' expectations and attitudes were assessed by three annual questionnaires. The students' opinions of and attitudes toward their classes were assessed by attitude scales and a semantic differential measure. Parents' opinions were assessed by a single questionnaire.

Student achievement was assessed by the Reading Comprehension, Vocabulary, Mathematics-Problem Solving, and Mathematics-Concepts scales of the Canadian Tests of Basic Skills, and by specially designed measures

of performance in Art and Composition. Students' academic self-concepts were assessed by the North York Self-Concept Inventory.

A specially designed observational system assessed teacher-pupil interaction, pupil participation, pupil satisfaction, method of instruction, subject emphasis, use of educational resources, physical characteristics of the classroom, and classroom atmosphere. Observations were also made with Indicators of Quality.

Different methods of statistical analysis were used with different variables. The unit of analysis for measures of parents' and teachers' opinions was the respondent; for all other measures it was the class. The final analysis included data from 62 classes -- sixteen of each of the two smaller sizes and fifteen of each of the two larger.

The teachers expected classes of 16 and 23 to have many advantages and classes of 30 and 37 to have many disadvantages. Their experience during the study confirmed most of their expectations. However, the other findings provided little support for these beliefs.

Varying class size produced few changes in classroom functioning, and most differences between class sizes were relatively small. A large significant difference observed was in the frequency with which pupils were addressed by teachers as individuals, which increased as class size was reduced below 30. However, teachers did not spend more time talking to individual pupils in classes of 16 or 23, so it seems that the differences observed were not the result of a change in classroom management, but simply the result of the same amount of individual attention being distributed among fewer pupils.

Other differences included: an increase in Mathematics Concepts scores in classes of 16, an increase in pupils' verbal participation in classes of 16 and 23, more frequent descriptions of classes of 16 or 23 as

quiet, more frequent descriptions of classes of 30 or 37 as having "lots of noise," an increase in Indicators of Quality scores in classes of 16 or 23, an increase in classes of 37 in the frequency with which pupils reported that not enough time was spent on Art. There also was some evidence that reading was taught more often in classes of 16 than in other classes. There were a few other differences which were identified: either they are not readily explained by class size differences or they offer further support for the above findings.

Caution must be exercised in generalizing the results of this study. They cannot be safely generalized to primary grades, high school or university. They cannot be generalized to classes operating on a rotary timetable, open plan schools, team teaching or split grades.

An important characteristic of this study which limits generalization of the results is that teachers were given no training in the use of appropriate methods to accomplish specified educational objectives with different class sizes. Essentially, this study changed class size and nothing else.

Acknowledgements

This study was possible only with the support and assistance of many people who have no expectation of recognition or specific acknowledgement -- the students, their parents, the staff of participating schools, especially those teachers who were directly involved in the study, and their principals and vice-principals. When the study was initiated, the research team promised the school anonymity and that we would treat all data from them or about the classrooms as confidential. Thus, it is not possible to recognize at all by name those who made the greatest contribution to the study, nor can we recognize another important group, the students, teachers and principals of several schools who made it possible for us to pilot test and refine many of the instruments used in the study.

We must also not overlook the support received from senior officials of various boards in Metropolitan Toronto and the Area Superintendents and their Assistants in the City of Toronto.

In one sense this study could be considered in part a tribute to the late Helen Sweet. As President of the Toronto Teachers' Federation (TTF) she presented the Board with the Federation's request that a well-designed study be undertaken to examine the effects of class size. Later on, as Past-President of TTF, she accepted responsibility for organizing and chairing a consulting group that was available to the research group as it explored and developed various ways of measuring the variety of class size effects. This committee, consisting of Jim Acheson, Denise Boudreau, Raymond Boyd, Elizabeth Clegg, Sandra Donaldson, Joyce Saunders, Sandra Statler and Mary Thomas, provided many helpful suggestions and comments.

Another large group of people are those from other institutions, especially The Ontario Institute for Studies in Education (OISE), who provided consultative assistance. Among the many people who provided valuable advice we would particularly like to thank Carol Attridge, Judith Dennison, Barr Greenfield, John Herbert, Tom Maguire, Les McLean, Jack Murray, Doris Ryan, Merle Wahlstrom, Dawn Whitmore, and Richard Wolfe. They were most helpful in developing instruments as well as procedures for data management and analysis.

The staff for this project also deserves a note of sincere appreciation. Dini Roper, as a Research Assistant at the start of this project, did much of the developmental work on the observation schedules. Important contributors to the study were the observers -- Vijay Agnew, Maralyn Biderman, Mary Jones, Shawn Moore, Barbara Passmore, Sheila Vierin and Kathy Wayne. Nor only did they carry out the classroom observations diligently and competently, but their suggestions and comments were extremely helpful in designing the observational instruments.

Among the Research Technicians special mention is due to Janis Carr-Harris for her contribution to the development of the rating scale for the Art measure, and to Frances Wells for her contribution to the development of the rating scale for the Composition measure.

Data management and programming for the study were facilitated by Mike Chovaz and Bill Postl. The preparation of numerous reports, instruments, and other documents would have been impossible without the typing skills and other assistance of Lynda Groves, Janice Hlywka and Karen Sommerville.

The acknowledgements of those people who provided valuable help would be incomplete if we did not also thank June Ball, Jocelyn Bates, Carole Brown, Lea Eistrat, Janis Gershman, Valerie Jordan, Janet Kirchmeyer, Gloria Maddever, Valerie McLeod, Joyce Moore, Sylvia Niit and Wendy Simmons for their help in completing a variety of tasks.

The Authors

Note to Reader

This report has four sections. The first two chapters, which describe how the study was designed and carried out, are quite long so that the reader may be made aware of the problems of conducting an experimental study of class size, the limitations of the present study due to these problems, and the efforts that were made by many people to reduce these problems to a minimum. The chapter entitled Methodology contains some descriptive data which some people might expect to find in the Results section. These data describe the sample, the instruments used for assessment, and the monitoring of class size.

The chapter entitled Results is relatively brief, although it may contain somewhat more explanatory material than is normal. The final chapter, Discussion and Conclusions, does not provide a summary of the procedures and findings, but examines the implications of the findings for the educational system and makes suggestions for further research.

A separate volume called the Technical Supplement contains more detailed information than this volume does about instruments, procedures and non-significant results. (Interested readers may obtain the Technical Supplement, on microfiche, from the Ontario Government Bookstore or, in printed form, on loan from the Ministry of Education Library.)

Contents

Abstract/*iii*

Acknowledgments/*vi*

Note to Reader/*viii*

INTRODUCTION/*1*

- History and Background of the Study/*1*
- Review of Literature/*4*
- Design Considerations and Development/*7*
- Scope and Limitations of the Study/*10*

METHODOLOGY/*14*

- Description of Sample/*14*
- Instrumentation/*19*
- Procedure and Design/*35*

RESULTS/*54*

- Teachers' Opinions and Attitudes/*54*
- Observation of Classroom Process Variables/*74*
- Student Opinions and Attitudes/*91*
- Student Outcome Measures/*101*
- Parental Opinions/*108*

DISCUSSION AND CONCLUSIONS/*117*

REFERENCES/*127*

Introduction

History and Background of the Study

The issue of class size has long attracted the interest of the educational community. A consistent theme in discussions of class size has been teachers' preference for small classes. One of many examples which can be cited is the 1972 poll of the National Educational Association in which the large number of students per teacher was ranked as the most "serious" or "critical" of 20 commonly cited teaching problems (Report on Educational Research, October 11, 1972).

It is important to note at the outset that the research on class size (which will be reviewed briefly in the next section) has not produced such consistent results as have surveys of teachers' opinions and attitudes. For example, Blake (1954) conducted an extensive review of the class size research. Upon analyzing studies that satisfied his established criteria of scientific adequacy, he found a mixed bag of results: some findings favoured small class sizes, others favoured large class sizes and still others showed no effects due to class size.

In the early 1970's several factors triggered a fresh round of debate on class size and pupil-teacher ratios in Toronto.* For one, the public was expressing concern about rising taxes due to the

---

* Both these terms (class size and pupil-teacher ratios) are commonly used, but the concepts are complex and not uniformly defined. Pupil-teacher ratio generally refers to a number of students divided by a total number of teaching and non-teaching educational staff (i.e. those without classroom teaching responsibilities). Class size refers to the total number of pupils in a classroom for instructional purposes. A discussion of these concepts illustrated with Ontario data, can be found in Ryan and Greenfield (1975).

high costs of education. Second, the Toronto school system was anticipating a decrease in student enrolment. In an attempt to curb increasing educational expenditures, the Ministry of Education introduced ceilings in 1971 which placed limits on school board spending in Ontario.* Since instructional personnel costs represent the largest portion of a school system's expenditure, there is a tendency for educational costs to vary inversely with class size and pupil-teacher ratios.

In light of the concerns and publicity surrounding "class size," a trustee of the Toronto Board of Education requested a report on research literature relating to optimum class size. In response, Shapson (1972) prepared a document which was issued as a regular research report entitled "Optimum Class Size? A Review of the Literature," (in November 1972) and subsequently reprinted in Ontario Education (May/June, 1973). The report highlighted the major issues surrounding research on class size (see following section).

The publication of this report resulted in further publicity, partly due to newspaper references. For example, in an editorial making reference to the report, the Toronto Star argued against the need to maintain "small" class sizes (November 18, 1972). In a second editorial (December 30th), which referred to the many letters which the previous article had generated, they continued to hold the position that the value of smaller classes was not proven. The Board had requested that the report be advertised for sale and the advertisements appeared in January.

The Toronto Teachers' Federation (TTF) wrote a letter to the Board of Education asking that a study on class size be undertaken.

---

* These ceilings subsequently were removed in 1976.

Their assembly had passed a motion "that TTF request the Board to institute a research project to determine the effects of class size on the education of a child in Toronto." This request was considered at the School Programs Committee meeting of February 20, 1973. In reply, the Committee made the following requests which were approved by the Board:

> "(a) that the Director of Research, through the Director of Education, be requested to submit a proposed research project, with costs, which might be undertaken in Toronto;
>
> (b) that the Ontario Institute for Studies in Education be requested to undertake a research project to determine the effects of class size on the education of a child;
>
> (c) that the Ontario Secondary School Teachers' Federation, District 15 and the Toronto Teachers' Federation be invited to become involved in any research study which may be undertaken."

A brief proposal was prepared to indicate the nature and magnitude of such an undertaking. The Committee also had a reply from TTF indicating their interest in co-operating in a study if they knew what role they might play, and a letter from the Director of the Ontario Institute for Studies in Education (OISE) indicating that the Institute was not in a financial position which would allow them to undertake and support such a study at this time.

After discussing the proposal, its implications and the cost estimates, the Board passed a motion directing that a proposal be submitted to the Ministry of Education for a research contract. In response to this motion, a letter was written to the Minister of Education enclosing the brief proposal which had been submitted to the Board. A favourable and encouraging reply was received and then

a more detailed proposal was prepared. The contract was agreed to and by the beginning of November, 1973 the project was officially underway.

Review of Literature

The strategy adopted in the review of the literature prepared by Shapson (1972) was to highlight the major issues surrounding the research on class size, rather than to conduct a detailed review of all available studies. The report presented evidence of inconclusive and inconsistent results, especially with respect to the effects of class size on student achievement, and it identified methodological weaknesses. It also attempted to illustrate that the class size question is a complex one which should not be dealt with in isolation from other important factors in education.

Because of the availability of the above report and because other thorough reviews have recently been published (e.g., Gajewsky and Shore, 1973; Lindbloom, 1970; Ryan and Greenfield, 1975) no attempt is made to repeat a detailed review of the class size literature. Rather, some of the major problems associated with research in this area are reviewed in this section.

The problem of conducting scientifically adequate research on class size is quite striking. In his extensive review of the literature on class size, Blake (1954) located 267 studies. He then proceeded to focus his review on those studies that used research procedures and was left with a sample of 85 studies. As a final step, Blake examined these 85 studies in terms of criteria of scientific adequacy. Only 22 studies satisfied these criteria.

The available research also indicates that other educational variables (e.g., the teacher, the subject matter and the method of instruction) are more important than class size. Because many studies have failed to control these variables, inconsistent results have been obtained and it has been difficult to assess the effects of class size itself.

Much of the previous research on class size is based on a correlational approach: the data were collected in existing situations with no attempt to set up experimentally different class sizes. Since many important variables have not been accounted for in the correlational studies, one cannot determine whether the results were attributable to class size or to a combination of variables (e.g., in some school systems it may be the custom to group together low achieving students in classes smaller than the average).

In the few cases where attempts were made to experimentally manipulate the size of classes, the studies were not well-designed (e.g., the large class was given to the most experienced teacher) and/or they included such a small number of classes that the results are readily attributable to teacher effects.

Another problem associated with class size research is that there has been a lack of agreement on the quantitative dimensions of large and small. A class size, 35 for example, that in one study is viewed as small (compared to size 50) turns out to be large in another study (compared to size 25).

Because of these methodological problems, it is not surprising that the research results have been inconclusive. This is especially true when student achievement is examined: some studies

have found that small classes produce better results, others that large classes are better and still others have found no "class size effect."

The most extensive studies involving class size were carried out at Columbia University (e.g., Olson, 1970; Vincent, 1968). Conducting observations in almost 20,000 classrooms, an assessment was made of the effects of a variety of variables on the "quality of education" in a school system.* The research was not concerned with the pupils' output (e.g., achievement).

The results of Olson's (1970) study demonstrated that class size was not the most important variable (subject being taught and method of instruction were more important). However, Indicators of Quality scores were found to be negatively correlated with class size; quality of education was found to be better in small classes than in large classes. Since Olson's study was not experimental, it is difficult to draw conclusions about class size effects; no information is provided about important variables which would have to be considered. For example, it is possible that the large proportion of smaller classes in the sample were from richer school districts. One would then want to consider whether richer districts had a greater ability to attract the best teachers, whether they had more instructional resources and whether they enrolled a different type of pupil than did the poorer districts in the sample. Such differences would make it difficult to

---

* While they were not initially intended to examine class size, the Columbia studies have been widely publicized in recent debates on class size. A review of the Columbia studies, focused on a critique and analysis of the instrument Indicators of Quality which was used to assess quality of education, is presented in the Technical Supplement.

establish the degree to which the results presented by Olson can be attributed directly to the effects of class size.

Previous research has also shown that although there is a strong expectation that smaller classes will result in more individualized instruction, teachers do not necessarily take advantage of this opportunity when their classes are small (e.g., Danowski, 1965).

The most consistent aspect of the class size literature is that teachers have well defined attitudes and feelings about the effects of class size; they express a strong preference for small classes. Students' perceptions on this subject remain to be explored.

Design Considerations and Development

The literature review which had been completed by Shapson (1972) helped to pinpoint the problems experienced in previous studies relating to class size and provided a frame of reference for developing the design of this study.

Based on the experiences of previous research, several decisions were made. First, it was decided that the study be experimental, involving some form of random assignment of teachers and pupils to the various class sizes. Also, some control over the assignments would have to be maintained throughout the study by the researchers and not by the school.

Second, because one characteristic of previous research was the lack of common definitions of "large" and "small," it was decided to include more than two different sizes in the study. This would ensure the representation of a broader spectrum of sizes in the study.

In addition, it was considered important to provide a rationale for those class sizes that were decided upon (cf. Procedure Section).

It was also desirable to run the study for two years with both students and teachers being randomly assigned to a different class each year. This was a method of further controlling the teacher variable. It was apparent from the beginning that administrative problems would make it impossible to do a huge study involving hundreds of classes. Thus, it was imperative to carry out the study in a manner that provided some balance in the design and to safeguard against the possibility that by chance alone the random assignments would provide an unbalanced distribution of teachers or pupils in the various sizes.

Another design feature concerned the decision to implement the study in the junior grades. As this appeared to be the first class size project with this degree of experimental manipulation, it was desirable to choose grades which were least likely to cause concern among staff or parents, and where the research undertaking itself would be most easily managed. In preliminary discussions, many people felt that class size was most important, educationally, in the primary division. This very attitude precluded running the study in the primary grades since many teachers and parents felt that children's progress would be unduly jeopardized if they were assigned to a large class.

Proposals to conduct the study in grades above the junior division would have caused less anxiety among parents and staff. However, rotary and semi-rotary organization for school subjects is

the general rule in these higher grades. It was felt that rotary organization presented design problems most notably with respect to teacher bias. For example, it might be possible for a teacher who saw several classes each day on a rotary basis to maintain a negative attitude toward the "large" class and a positive attitude toward the "small" class throughout the year. Our opinion was that it would not be nearly as easy for a teacher to maintain a particular bias in the junior grades where the same teacher meets with the same pupils every day, all year. While a teacher initially might think that he could treat the students differently because of his attitude toward teaching a small (or large) class, the necessity of discharging his responsibility to the same pupils every day for the whole year would likely reduce or even remove the effect of any such initial bias.

Students in the junior grades are able to handle simple written materials: paper and pencil tests could be used to collect information from them. The testing of children in the primary grades presents many problems and might deliver less valid results.

The final considerations related to the kinds of effects that would be measured. From the beginning it was assumed that class size might make genuine and noticeable differences in an area that might loosely be referred to as "classroom process," a collection of variables (e.g., teaching method, pupil-teacher interaction) which could best be assessed through direct observations of classrooms. For example, Ryan and Greenfield (1975) concluded that reducing class size may produce increases in academic achievement only if teachers take advantage of the greater opportunity provided by smaller classes

to individualize instruction. Therefore, it follows that it should be important in a study of the effects of class size to record teaching methods and this is most efficiently done by direct observation.

A general rationale for examining process variables has been provided by Vincent and Olson (1972) who contend that a "good" school is expected to do more than teach its pupils to get high marks on achievement tests. As they point out in the following analogy:

> ... we base our evaluation of a place to live
> not on the number of persons it will accommodate,
> but the manner of life that can be lived there. (P. 4)

Similarly, in studying schools, we should be interested not only in their pupils' performance on achievement tests, but also in the conditions under which pupils learn.

It was decided, however, that some form of achievement testing would have to be incorporated because it would be a topic about which questions would be asked when the study was completed. For the same reason, the whole area of opinions and attitudes had to be incorporated in the design. With students this represented new ground because, while the opinions and attitudes of teachers toward class size have been well documented and studied in previous research, those of students have been ignored.

Scope and Limitations of the Study

It is important to note that this undertaking is only one of many possible studies of the effects of class size. The Ministry of Education simultaneously signed a second contract with OISE to clarify research questions and to plan different kinds of studies because it

recognized the complexity involved in planning and conducting studies of the effects of class size. In the report of OISE's contract, Ryan and Greenfield (1975) suggest that five different types of research are needed on class size and related problems. Within Ryan and Greenfield's overall scheme, this study tackles the problem from one perspective, as an analytical research question.

This study is limited to an examination of the effects of class size which are measurable within the framework of a traditional classroom organization, that is to say, the effects of varying the number of pupils taught in a classroom by one teacher. The study was not designed to examine changes in other forms of school organizations. Therefore, a consideration of which organizational variables most affect what goes on in the classroom is beyond the scope of this study; questions such as whether it is better to utilize resources for educational aids or for educational services or for altered class size are not considered. In addition, there was no attempt to examine class size from a flexible organizational viewpoint tailored to fit unique or particular needs. Only a few previous studies have considered flexibility in class size policy. Balow (1967), for example, reported an experiment in which a 50 per cent reduction in class size was accomplished for reading instruction by the simple expedient of staggering the departure and arrival times of students.

Obviously, the class size question has important financial considerations: decreasing class sizes, with all other factors held constant, is necessarily an expensive proposition. An examination of the financial implications related to changing class size is also out-

side the scope of this study as is a cost effectiveness analysis to compare the relative effectiveness of alternative components of school expenditures. There is no attempt in this study to examine different options for the same expenditure.

In discussing the relationship between class size and the method of instruction, the following opinion has been offered:

> Any criterion employed to assess the effect of class size is in actuality assessing the accomplishments of some method -- the method of teaching which was used in the study in question. Whether it appears to better advantage in large classes or in small classes depends upon the compatibility of the method with the size of the groups being investigated. (Encyclopedia of Educational Research, 1969, p. 142)

However, research in this area is further limited by the fact that there is no consensus (either from professional educators or the public at large) concerning what the goals, objectives and outcomes of the educational system should be. As far back as 1932, Carr stated that before further work on the effects of class size be undertaken, the purpose of elementary school education should be clarified in order to give direction to subsequent investigation.

This state of affairs obviously has implications for any research design: it means that in the present study teachers were not given a set of objectives for each class size, nor were they encouraged to change their procedures, styles or teaching methods, although participating teachers obviously had their own expectations of the effects of class size to guide themselves (see Results for data on teachers' expectations prior to the study). Thus, it is important to note that this study does not explore what could be done with each class size, but rather what actually is done.

If it were possible to reach consensus in specifying objectives for each class size, then an alternate and feasible way of exploring the class size question might proceed as follows:

-- the research design could specify a method most compatible to the objectives for each class size;

-- training in these methods could be provided;

-- one could determine whether the method specified was actually consistently employed;

-- one could determine if the specified objectives were actually met in each class size;

-- one could conduct a cost effectiveness analysis.

Unfortunately, this approach still appears to be a long way off.

Methodology

Description of Sample

Description of Schools

Twelve schools participated in the study, eight from the City of Toronto and two each from the Boroughs of North York and Scarborough. Each of the six administrative areas in the City of Toronto is represented in this sample. Some demographic characteristics of participating schools are presented in Table 1.

TABLE 1

SOME CHARACTERISTICS OF SCHOOLS PARTICIPATING IN THE CLASS SIZE STUDY*

Location of School	% of Households with Income under $10,000	% of Dwellings Owner Occupied	% of Population Speaking English as Mother Tongue
Toronto	60	70	68
Toronto	65	64	51
Toronto	77	42	75
Toronto	61	43	59
Toronto	62	52	43
Toronto	62	53	41
Toronto	46	71	48
Toronto	36	60	91
North York	61	13	56
North York	86	3	64
Scarborough	46	54	73
Scarborough	37	73	84

* Based on 1971 census data by matching school boundaries as closely as possible to census tracts.

Description of Teachers

In the first year of the study the sample consisted of 34 teachers. In all participating schools efforts were made to ensure that teachers were given alternatives regarding whether or not they would participate (see Procedure Section for more details). Two other criteria considered in the selection process were that the teachers be prepared to give a two year commitment to the study (i.e. they were not planning to leave teaching or transfer schools) and that they have at least two years of teaching experience. Of the 34 teachers, only three did not participate in the second year of the study.* Replacements were made as necessary for the second year (see Table 12).

Slightly over 70 per cent of the participating teachers were women (this coincides with the system-wide figure for elementary school teachers). Fifty per cent of the teachers in the study had five years or less of previous teaching experience (with the first year's sample, the range was from two to nineteen years experience with a mean of 7.2 years).

Approximately 80 per cent of the teachers had previous experience in Grade 4 or Grade 5 (with 35 per cent having experience in both these grades). In addition, 70 per cent of the sample had previous experience in the primary grades, with 27 per cent having taught in Grades 6 to 8.

In the year prior to the study, 53 per cent of the sample taught Grade 4 and the mean class size for all participating teachers was 30.97 (ranging from 18 to 38).

---

* Of those teachers leaving the study, one went on maternity leave, the second was given an opportunity for professional advancement, and the third left because of a strong personal preference not to change grades (a condition of the second year of the study).

When asked what was the smallest class size they had ever taught, 41 per cent of the sample indicated experience in sizes of 25 or smaller (the smallest size mentioned was 5). With regard to the largest class, all participating teachers had taught in sizes of 30 or larger (the largest size mentioned was 49).

Description of Students*

The sample for the first year of the study consisted of Grade 4 pupils in participating schools. These pupils moved to Grade 5 for the second year of the study (see Procedure Section for more details).

Information concerning participating students' country of birth is presented in Table 2. Approximately 72.9 per cent of the students were born in Canada; this matches the figure for the total Toronto elementary school population (70.1 per cent). About half (49.4 per cent) of the students in the sample reported learning English as their first language. For those students whose mother tongue was not English, the main language groups represented in the study are Greek, Italian and Chinese, followed by Portuguese, Ukrainian and German.

Information on socio-economic levels for heads of households for participating students is presented in Table 3; 52.2 per cent of the sample represent the lower socio-economic levels (Categories 25-31.99; Blishen, 1967) while only 8.2 per cent represent the higher socio-economic levels (Blishen's Categories 55-76.99). The comparable figures for the total Toronto elementary school population are 43.8 per cent and 13.2 per cent respectively.

---

* The descriptive background information discussed in this section was available from the 1975 Every Student Survey (Deosaran, Wright and Kane, 1976) for participating Toronto students. Similar data were collected for participating students from other Boroughs as part of the Class Size Study.

TABLE 2

COUNTRY OF BIRTH FOR PARTICIPATING STUDENTS

Category	Elementary School Population (1975)* %	Class Size Study Sample** %	(n)	Range*** (Class Size Study) %
Born in Canada	70.1	72.9	(525)	54.0 - 94.7
- English 1st Language		(41.2)		(8.5 - 85.7)
- English not 1st Language		(31.6)		(6.1 - 52.5)
Not Born in Canada	29.9	27.1	(196)	5.3 - 45.9
- English 1st Language		(8.2)		(3.2 - 16.0)
- English not 1st Language		(19.0)		(0.0 - 36.9)

* Based on data for the elementary school population in Toronto reported by Deosaran, Wright & Kane (1976).

** Toronto schools only.

*** The range is based on data for individual schools participating in the study.

Tables 2 and 3 are important to consider in generalizing the findings. The population studied is drawn from a large urban area. While this includes a wide variety of people, and thus can be seen as broad-based, it does not represent the relatively homogeneous population, in terms of these variables, which might be found in some suburban or rural communities.

TABLE 3

SOCIO-ECONOMIC LEVELS FOR HEADS OF HOUSEHOLDS
FOR PARTICIPATING STUDENTS

Blishen's Category*	Category Description	Elementary School Population (1975)** %	Class Size Study*** %
25-31.99	Labourers, Truck Drivers, Porters	43.8	52.2
32-38.99	Bartenders, Sheetmetal Workers, Repairmen, Sales Clerks, Jewellers, Stationary Engineers, Machinists	9.5	10.5
39-42.99	Pressmen, Printing Workers, Electricians, Members of Armed Forces, Clerical Occupations	8.5	7.9
43-54.99	Actors, Tool & Diemakers, Medical and Dental Technicians, Embalmers, Real Estate Salesmen, Musicians, Stenographers, Athletes	10.8	9.7
55-76.99	Clergymen, Various Owners and Managers, Insurance Salesmen, Librarians, Teachers, Professional Engineers, Physicians, Computer Programmers, Air Pilots	13.2	8.2
---	Pensioner, Retired, Workman's Compensation, Disabled or Ill, Welfare, Mother's Allowance, Adult Training or Retraining, Unemployed, Mother Only, Housewife, Respondent on His/Her Own, Group Home Head (e.g., Social Worker)	10.5	8.0
---	No Information or Unknown	3.7	3.5
	NUMBER OF CASES	(61,688)	(722)

* These categories are based on Blishen's (1967) socio-economic index for occupations in Canada.

** Based on data for the elementary school population in Toronto reported by Deosaran, Wright & Kane (1976).

*** Toronto schools only.

Instrumentation

The Toronto Classroom Observation Schedule

Selection of Variables and Procedures for Observation. The selection of variables for observation was made difficult by the absence of any scientific theory of class size effects and the lack of agreement as to what constitutes "good" classroom functioning or the "quality of education."

Consideration of classroom functioning and the existing literature on class size (Sitkei, 1968; Shapson, 1972) produced the following list of selected areas to investigate through classroom observation: teacher-pupil interaction; student participation; student satisfaction; method of instruction; subject emphasis; physical conditions; use of educational resources; and classroom atmosphere.

The Toronto Teachers' Federation (TTF) resource group* was asked to comment on the list of variables and on procedures for observing classrooms. The group's main concern was that the observational data be "representative." They felt that representative data could not be obtained through brief observations. Consequently, it was decided that classes would be observed half a day at a time, eight times throughout each year.

Four other considerations were also important in the development of the observation instrument. First, the instrument had to be comprehensive enough to assess effects in the selected areas. Second, it had to be suitable for use in functioning classrooms; that is, it

---

* Information concerning this group is provided in the following section (Procedure & Design) although it is referred to briefly with respect to the development of several instruments in this section.

should be as unobtrusive as possible and easy to administer. Third, in light of contradictory findings about the effects of class size, it was considered important to have as objective an instrument as possible, capable of providing reliable assessments of classroom processes. Finally, the instrument had to be appropriate for classes in Ontario schools.

An examination of many previously used observation systems reveals that while many satisfied several of the considerations, none satisfied all of them.

Some of the observation systems considered -- for example, Herbert's System for Analyzing Lessons (1967), or Bales' Interaction Process Analysis (1950) -- were simply not practicable, requiring audio-visual equipment, extensive data analysis, and more than one observer to code a single observation.

Others were not suitable for use in a junior grade, or required pupil ratings of lessons and teachers. For example, Cogan's Measurement for Teacher-Pupil Interaction (1958) was administered in junior high schools where formal lessons could be taught repeatedly to different classes. Collection of data involved a complex rating of teacher behaviour by pupils.

Other systems appeared to be too narrow in scope for the present study, assessing only a few aspects of the classroom; for example, Flanders' Interaction Analysis (1970) and Shapiro's Observation of Classroom Behaviour (1963).

The Columbia University instrument, Indicators of Quality (Vincent & Olson, 1972), although comprehensive and unobtrusive, apparently does not assess subtle variation in the degree of effects in the selected areas. As well, it was intended for use in the evaluation of school

systems rather than of classrooms. Despite these limitations, it was used as a supplementary instrument (see the section, Indicators of Quality, to follow).

Either many separate instruments would have to be used, or a single measure meeting all five criteria would have to be devised. It was decided to develop locally a single observation schedule to satisfy the requirements.

Instrument Development. During the planning year before implementing the study in Grade 4, the Toronto Classroom Observation Schedule (TCOS) was designed. Observers were hired and trained to use the instrument, which was then field tested in the Spring of 1974 in classes of teachers who would be participating in the study. As a result of the field testing, the section of the TCOS intended to assess classroom atmosphere was replaced and some revisions were made in other sections (because of these changes, data from the field test could not be used as pre-measures).

The revised instrument was field tested in September, 1974, in non-participating classes, and was found to be adequately reliable for use in the study. A detailed description of the field testing and observer training is provided in the Technical Supplement with reliability data.

In the revised version of the TCOS, the following variables are investigated:

(1) Teacher-Pupil Interaction. In order to investigate these effects it was decided (after Flanders, 1970) that a useful variable for observation would be the verbal behaviour of the teachers. It was further decided that three aspects of teacher speech would be most likely to be sensitive to variation in class size. These are:

(i) content -- that is, whether a verbalization is related to pupils' behaviour, course content or routine classroom procedures;

(ii) function -- the function of an utterance can be categorized as approving, disapproving or neutral. Neutral verbalizations about course content can be further categorized as questioning or telling;

(iii) audience -- whether an utterance is addressed to an individual, a group, the whole class or other adults.

Data on teacher verbal behaviour are collected during an eight minute period. Every fifteen seconds during the eight minutes the observer categorizes the teacher's utterance occurring at that time.

(2) Student Participation. Student participation is measured in two ways, by observing both the whole class and randomly selected individual pupils. In one part of the observation schedule counts are taken at thirty second intervals (for three minutes) of all the students in the class who are, and are not, performing the assigned task, or who have not been assigned a task. In another section, the participation of individual pupils is observed in more detail. The observer classifies participation as verbal or non-verbal. If participation is verbal, it is further classified according to whether it is self-initiated or reactive, and by target audience (i.e. teacher, peer or self). The pupil can also be classified as off task or as having no task. Data are collected at fifteen second intervals for four minutes per individual.

(3) Pupil Satisfaction. Cannon (1966), in a study of "large" and "small" kindergarten classes, reported that aggressive acts were more frequent in the larger class and that in the larger class "a high level of frustration appeared to advance the general climate of aggressive behaviour" (page 10).

While Cannon's study included only two classes and was not well-designed, and while no evidence of the "high level of frustration" is offered, it was decided that observations of aggressiveness in the class size study would permit not only a test of the validity of this finding, but also an assessment of pupil satisfaction. The frequency of aggressive acts would theoretically be negatively correlated with pupil satisfaction. A sign system of observation, as defined by Medley and Mitzel (1963), is used to record aggressive acts. A sign system provides a record of the frequency of specified acts or incidents - all overt, disruptive, hostile acts occurring within a five minute observation period. The number of different pupils acting aggressively is also recorded.

(4) Method of Instruction. The observer ranks by frequency the methods of instruction used by the teacher during the observation period.

(5) Subject Emphasis. The subjects taught during an observation visit are recorded by the observer and ranked according to the amount of time spent on them.

(6) Physical Conditions. Classroom noise, furniture arrangement and educational aids are noted by the observer on a series of checklists.

(7) Use of Educational Aids. Estimates are made of the proportion of pupils using audio-visual, written and mechanical or concrete aids.

(8) Classroom Atmosphere. The observer makes ratings on Likert items* based on aspects of classroom functioning considered representative to classroom atmosphere (for example, the pupils' regard for the teacher).

The sequence of activities during a half-day classroom visit is presented in Table 4. A copy of the TCOS and the manual for its use can be found in the Technical Supplement.

Indicators of Quality

Indicators of Quality was developed in the early 1960's by William S. Vincent and colleagues at the Institute of Administrative Research, Teachers College, Columbia University. It is an observation instrument intended to measure "quality of education," a concept comprising four categories of classroom activity: individualization, interpersonal regard, creative expression and group activity (see Vincent & Olson, 1972).

The instrument consists of 51 items, each of which describes one "positive" and one "negative" classroom activity (or "sign"). An observer spends approximately twenty minutes in a classroom and for each item indicates whether what was observed can be described by either the

* Scales in which observers are asked to indicate the degree to which a statement describes an aspect of classroom functioning.

TABLE 4

TORONTO CLASSROOM OBSERVATION SCHEDULE:
SEQUENCE OF ACTIVITIES DURING A HALF-DAY CLASSROOM VISIT

Arrival - identifying individual pupils (with teacher's help) for "Individual Pupil Participation"

Systematic Observations - (fixed time samples)

Round I - 24 minutes

Activity	Total Time	Unit Time
Teacher Verbal Behaviour	8 mins.	- sampling teacher verbalizations every 15 secs.
Individual Pupil Participation	8 mins. (4 mins.for pupil x) (4 mins.for pupil y)	- sampling participation of one pupil every 15 secs.
Pupil Participation	3 mins.	- sampling participation every 30 secs.
Pupil Aggressive Behaviour	5 mins.	- recording as occurs within 5 minute period

Round II - repeat Round I activities (24 minutes)

Round III* - repeat Round I activities (24 minutes)

Observation Checklist - 20 minutes

(e.g., method of instruction, subject emphasis, use of educational aids and physical conditions)

Classroom Atmosphere Ratings - completed during the last five minutes of the visit.

* During morning visits, three rounds are completed, while two rounds are completed in the afternoon.

positive or the negative sign. (If neither of these signs is observed, an item is not scored.) The most commonly used scores are the number of positive signs, the number of negative signs, and the difference between the numbers of positive and negative signs.

Indicators of Quality became important in the discussion of the class size issue after Olson (1971) published a study which showed that Indicators of Quality scores increased as class size decreased. While Olson's study was not experimental and did not demonstrate that class size directly affected Indicators of Quality scores (it might simply have been correlated with another variable which did affect the scores), it was widely publicized and has been cited as evidence that reducing class size will improve "quality of education" (for example, by the British Columbia Teachers' Federation, 1973).

Indicators of Quality has also been subjected to much criticism (for example, from the British Columbia School Trustees' Association, 1974; Ryan and Greenfield, 1975). While not all of the published criticisms are relevant or valid, it is true that the test's designers have provided little evidence of the test's validity or reliability. As part of this study, a detailed critique and statistical analysis of the instrument was prepared (see the Technical Supplement). The reliability data collected in this study proved the instrument to be satisfactory as it was used in this study.

Despite its shortcomings, Indicators of Quality was included in the battery of instruments selected for the study. There were several reasons for this decision. Perhaps the most important reason was that Olson's (1971) data suggested that the instrument was measuring some important aspect of classroom functioning. In spite of there being no consensus about what constitutes "quality of education," the decision to

include the instrument in the study was reinforced by the extensive consultation with educational experts which was undertaken when the test was being developed. Two hundred educational experts were asked to describe the classroom activities they considered typical of a good school; these descriptions were then used as guides in the construction of items on the scale. Since the experts consulted ranged from Carl Rogers to B. F. Skinner, the behaviours selected for observation should be relevant to almost every branch of educational theory.

Second, although Olson's (1971) study was not experimental, Indicators of Quality scores demonstrated a clear relationship to class size. For this reason, it seemed important to include the measure in an experimental investigation of the effects of class size.

As well, Indicators of Quality devotes more attention to certain areas (for example, individualization and creativity) than the TCOS does. Finally, at the time decisions about instrumentation were being made, the TCOS was in its developmental stage and so it was not known how valid and reliable the TCOS would prove to be.

The Canadian Tests of Basic Skills

The following four subscales of the Canadian Tests of Basic Skills (CTBS) Form 1, (Thomas Nelson & Sons, Canada Ltd., 1968) were selected as standardized measures of student achievement:

> Vocabulary (17 minutes): Tests general skills in the use of tools involved in word recognition, knowledge of the meaning of words, and sensitivity to fine differences in meaning and judgment in choosing the most appropriate word in a given contextual situation;
>
> Reading Comprehension (55 minutes): Tests skills in recognizing and understanding details and relationships, in discerning the purpose or main idea of a selection, in organizing ideas, and in evaluating what is read;
>
> Mathematics Concepts (30 minutes): Tests skills in understanding the number system, mathematical terms and operations;

Mathematics Problem Solving (30 minutes): Tests skills in solving mathematical problems.

The tests were administered during the Spring for three consecutive years: once to obtain a pre-measure in the Spring prior to implementing the study (Grade 3 level of the test), once toward the end of the first year of study (Grade 4 level) and finally toward the end of the second year of study (Grade 5 level).

Research technicians were hired and trained to administer the tests. The standard administration manual was followed except for two modifications:

(a) it was discovered that the instructions provided by the publisher were too formal for Grade 3 pupils who generally had no prior experience with standardized testing procedures. Consequently, a reworded version of the publisher's instructions was prepared adding information concerning finishing early, etc., and a new procedure for identifying answer sheets;

(b) after considering the length of the Reading Comprehension Test, (55 minutes), the concentration span of pupils and the scheduling of the recess in some schools, pupils were allowed a break for recess instead of the normal "brief stand-up time" prescribed.

Art Sample

The rationale for including a measure of students' art development was provided in a meeting with the TTF resource group where concerns were expressed that standardized achievement tests focused on "core" subject material and therefore could not expose class size effects in pupils' development in non-core subjects, such as Art. The process of developing a methodology for analyzing children's art was aided by previous experience of the Research Department; Pennington's (1967) Research Report #43 was a major theoretical source:

> The literature firmly established developmental stages, relying upon the criterion of increasing naturalistic representation: the child's early tendency to motor expression and free representation of feeling, percepts and concepts increasingly gives way to visual realism. (P. 49)

Developmental stages in art are described in Children and Their Art, by Gaitskell (1970) who had considerable experience with elementary school art curricula in Ontario. Based on Gaitskell's work, the following five categories for rating the art sample were specified:

(1) Manipulative Stage. The child is exploring and experimenting with materials and form, beginning with a few "scribbles" and ending with a name given to a shape after it is drawn.

The Symbolic Stage, in which the child is able to produce symbols for his ideas by giving shapes desired characteristics, was subdivided into three categories: primitive, middle and advanced.

(2) Symbolic Primitive. The symbols have a suggestion of basic outline or shape. There is a definite lack of internal detail.

The objects appear to be placed on the paper according to whim; therefore, the spatial relationship is poor. There may be a tenuous baseline.

Colours seem chosen without regard to realism frequently enough that it is obvious through the picture.

(3) Symbolic Middle. What the symbols represent is clear, but the objects are still basically outlines. The basic shapes "houses" and "trees" that are so often used to represent children's art are evident.

Some detail in clothes and features appears, but it is not consistent. There is much copying of figures at this stage. Figures tend to be actionless. (Do not confuse with the wooden, but detailed, figures of the advanced stage.)

Object placement has a relationship to the situation being presented.

There is either a single baseline or definite multiple baselines.

Colour is not necessarily true to life.

There is often a "dead" feeling, a "meagre talent" feeling, and a "pretty pattern" feeling.

(4) Symbolic Advanced. The figures are definitely symbols, but are drawn with skill and with attention to internal detail. The details include patterned clothing, rosy cheeks, eyebrows, elongated noses, trees with branches and roots.

Figures may have good action, but some may be very detailed, yet still wooden.

Object placement may indicate a beginning realistic attempt at perspective or there may still be base-lines.

There is some evidence of colour muting or texture, but the feeling is that they were good ideas copied from neighbours, rather than an attempt to draw from life.

There is a sense of drawing sophistication, position sophistication, a finished picture.

The student seems to be trying to produce a work of art and the viewer gets the impression of some talent.

(5) Preadolescent. The picture shows an obvious attempt to photograph objects and their environment as they actually appear to the eye. A few symbols may still exist in some drawings.

Judge figures, objects, colour, texture, perspective for realism. If there is a serious attempt at photographic effect in any of these areas, rate the drawing "5."

Cartoon figures are a valid art form in picture making. If the characters are sophisticated and the drawing tells a story, rate the drawing "5."

The preparation of the instructions to teachers and pupils and the selection of a suitable topic for the pictures ("What I Like to Do") was facilitated by the suggestions of staff in the Art Department. One specialist in Art Education and one researcher were employed to rate the art sample on a five-point scale using the above categories. Agreement in rating, estimated by the Pearson product-moment correlation, was .78 (first year) and .90 (second year). Further details of this procedure and examples of students' art work are presented in the Technical Supplement.

Composition Sample

As with art, the decision to collect a sample of student compositions was an attempt to investigate an area of student achievement which was not measured by standardized achievement tests. A rating scale for assessing the compositions was developed from a similar procedure

previously used by the Research Department (Wright & Reich, 1972), after examining a sample of the compositions received. The compositions were rated according to five categories, having the following general characteristics:

1) Illegible. A good many papers in this group are illegible. Sentences are sometimes meaningless. Poor spelling and grammar often add to the difficulty in reading or understanding what is written. Continuity is lacking and generally stories are not very interesting.

2) Poor. These stories are for the most part not very imaginatively written. They are too often colourless, devoid of style and limited in vocabulary. Sentence structure is often poor and in some cases non-existent; the stories themselves are on occasion incoherent and lacking in form. The mechanics leave much to be desired.

3) Average. The compositions in this group are written well enough to be readily understood. Some do not have the colour and imagery evident in the following two categories, nor is the vocabulary generally as rich. The stories themselves and the sentences therein are usually complete. The mechanics are quite acceptable, enabling the stories to be read fluently.

4) Good. These stories are interestingly written with imagination and colour. Sentences are well constructed and story forms are logical and complete. The students usually spell accurately and appear skilled in grammar.

5) Superior. These compositions are originally presented, imaginative in style and descriptive in vocabulary. They arouse and hold the reader's attention throughout. The writers use dialogue advantageously, giving the compositions added interest and vitality. Sentences are well formed, varied and coherent. The stories progress in a logical fashion and mechanics seem well established.

The topics selected for students to write about were "Dreams" (first year) and "Wishes" (second year). Standard instructions for teachers and pupils were prepared and two research technicians who had teaching experience were employed to rate the compositions.

The process of grouping the compositions into categories was complex since features of two categories often were present in one story (e.g., a descriptive and interesting tale hampered by an absence of basic mechanics and legible handwriting). Nevertheless, rater agreement was high, .72 (the first year) and .89 (the second year) as estimated by the Pearson product-moment correlation coefficient.

Further details of this procedure and examples of students' compositions are presented in the Technical Supplement.

Self-Concept

There has been much discussion recently about the schools' role in enhancing a student's self-concept. Edeburn and Landry (1976), for example, point out that:

> If in the elementary school cognitive growth is emphasized and affective growth ignored, the child may be subjected to a highly competitive weeding process that could seriously erode his feelings of personal adequacy. (P. 372)

Purkey (1970) provides a good review of the many different definitions and dimensions of self-concept.

The North York Self-Concept Inventory developed by Shapson, Virgin and Crawford (1971) was selected to measure the self-concept of pupils with respect to their school environment. The Inventory was constructed from an analysis of three previously existing scales administered to North York students and the criteria for selecting items were that their content deal with either "scholastic" or "general self" dimensions of self-concept (dimensions dealing with "family" or "peer relations" for example, were excluded) and that each item discriminate between groups of high- and low-scoring students.

Since its development, the instrument has been used extensively in North York; the test/re-test reliability co-efficient is .81, indicating highly consistent response patterns.

The inventory contains thirty items. Students are asked to indicate whether they feel each statement is "true" or "not true" for them. Each item is scored by assigning one point for the appropriate response indicating a positive self-concept. A copy of the instrument is provided in the Technical Supplement.

Pupils' Attitudes toward School Questionnaire

Although the attitudes and feelings of elementary students on the effects of class size had been virtually neglected in previous research (cf. Shapson, 1972), they have been extensively investigated independent of the class size question. In developing this questionnaire, many sources were initially reviewed (Anderson, 1970; Anderson and Walberg, 1972; Beere, 1973; Coleman, 1966; Fox, 1966; Hall, 1966; Harris, 1966; Lorber, 1969; Steele, 1971; Stephens and Berryman, 1968; Stones, 1969; Wahlstrom, 1971; Walberg and Ahlgren, 1970; Yamamoto, Thomas and Karns, 1969). Based on these sources a list of items which might be related to class size was constructed.

In meetings with the TTF resource group, these items were reviewed and several additional items and alternative forms of items were generated. From these items, two preliminary forms (A and B) of the questionnaire were developed; on each form approximately half the items were common. The two forms were then pilot tested in Grade 4 classes in schools not participating in the study.

In constructing the final version of the questionnaire, each item had to be significantly correlated ($p < .05$) with the total score

and its response pattern was examined to see if it displayed any peculiarities such as a high rate of neutral response in all classes used in the pilot testing. Using these criteria, the final version contained 36 items which could be administered in less than one-half hour of class time. The items in the instrument were selected to probe for differences between class sizes with respect to pupils' happiness and satisfaction, reaction to the physical environment, reaction to the social environment, attitudes toward teachers and other pupils, and toward subject taught.

A copy of this questionnaire and further details of its development are presented in the Technical Supplement.

Teachers' Questionnaire

Two questionnaires for teachers were developed. The first was a background information questionnaire administered after the teachers had volunteered to participate in the study, but before they were informed of the class size they would be assigned to for the first year of the study. This information was used to:

(1) describe the participating teachers, e.g., previous experience, teaching preferences;

(2) sample some of the feelings and attitudes of teachers before they experienced the class sizes in the study;

(3) enable teachers in the study to express their expectations of the effects of changing the size of their class.

The second questionnaire, developed in co-operation with the TTF resource group, was designed to survey a wide range of teachers' attitudes, perceptions and opinions toward their assigned class size, and was administered in both years of the study. Some questions from the Background Information Questionnaire were included to check for changes in attitudes following experience with a particular class size.

The responses to this questionnaire were used to answer the following questions:

(1) to what extent are teachers' expectations concerning the effects of changing class size confirmed or denied by their experience with that class size?

(2) what, in teachers' opinions, are the major effects of class size?

A copy of both questionnaires and descriptive data are provided in the Technical Supplement.

Semantic Differential

A semantic differential was selected as a supplementary technique of assessing the attitudes of students and teachers toward school. This technique involves rating a concept, such as "My Classroom" according to pairs of polar adjectives (e.g., good...bad) which are separated by a five or seven point scale.

The semantic differential literature was reviewed (e.g., Kirby and Gardner,1972; Oliva, Maguire and Pacey, 1971; Osgood, Suci and Tannenbaum, 1961) and four sets of polar adjective pairs were selected to describe pupil attitudes toward the concept "My Classroom." These items were discussed with Professor T. O. Maguire who had used a similar instrument to evaluate individualized instruction programs in Alberta. The instrument was pilot tested in schools not participating in the study; one pair of adjectives which did not significantly correlate with the total score was dropped and the order of presenting the adjective pairs was modified.

Two forms of the semantic differential were developed for participating teachers: one to describe "My Classroom," the second to describe "The Pupils I Teach." A copy of the forms that were used and descriptive data are presented in the Technical Supplement.

Parent Questionnaire

The Parent Questionnaire was used to survey parental opinions and feelings about class size and its possible effects on their children. Principals of participating schools and the TTF resource group provided suggestions that helped in its development. The questionnaire was sent to parents toward the end of the second year of the study, to be returned in postage-paid envelopes to the Research Department. A copy of the questionnaire, which was also translated into Italian, Greek, Chinese and Portuguese, is provided in the Technical Supplement.

Procedure & Design

This part of the report is presented in five sections which:

(a) describe how details of the design were developed during the implementation of the project;

(b) describe the procedures that were followed for obtaining school participation;

(c) describe the assignment of students and teachers to class size;

(d) describe the assignment of class sizes to schools and the subsequent monitoring of these sizes;

(e) discuss the scheduling of data collection activities in the schools.

While the five main features are presented in text, the Technical Supplement provides a more detailed report of the procedures, especially for sections (b) and (e) above.

Implementing Details of the Design

TTF Resource Group. Because the Toronto Teachers' Federation (TTF) originally asked the Toronto Board of Education to undertake a class size study, the Federation was contacted after the contract was signed and a resource group of eight members, with various positions in the school system, was organized to meet with the research team. The

resource group made important contributions to the development of the research undertaking. Comments on draft materials helped to refine procedures, improve instruments and broaden the study by emphasizing classroom process variables such as the way that classes were conducted and how students behaved. The observation schedule which was developed reflected the resource group's strongly expressed belief that brief observation schedules would not adequately represent classroom activity. The group also strongly suggested the inclusion of measures of students' performance in areas such as art and composition to supplement the type of performance assessed by standardized achievement tests.

Selection of Class Sizes. Alternative class sizes were considered at a meeting of principal officials of the Toronto Board of Education, and it was there that the sizes finally adopted were suggested: 16, 23, 30 and 37, with a variation of ±2 students allowed.

This set of four class sizes is attractive for several reasons. The four sizes were spaced to permit some variation due to transfers. The size 40, which has very negative connotations, is avoided. The largest class size, however, would be larger than 35. This size is found sometimes in Toronto elementary classes, although administrators call it "large" and "more than desirable." Class size 30 appears to be a common size.* On the other hand, it was felt that teachers saw a class size of 25 or smaller as ideal, and the Columbia Teachers College studies on "quality of education" suggest that 25 is a size below which quality increases. The smallest class size was debated at some length; the feeling was that a class must have fewer than 20

---

* In the year prior to implementing the study, the average class size for all participating teachers was 30.97.

pupils to be "small." It seemed that the study would appear incomplete if a size this small were not included.

Staffing Provisions and the Involvement of Other Metro Boroughs. Each set of four class sizes requires approximately 106 students and four teachers, although the staffing formula provided only 3.5 teachers for that number of students. The contract did not allow funds to be used for extra teaching staff. Staffing relief for the participating Toronto schools was provided by assigning a small number of teachers to this project.

Because the number of participating Toronto classes was relatively small, and also in order to provide a broader base for the study, the school boards from the other Metro boroughs were invited to participate. A presentation was made to the researchers from the other boroughs, Federation representatives in the boroughs were contacted, and support was formally obtained from the Council of Directors. With this co-operation two schools from each of Scarborough and North York volunteered to participate.

Selection of Grade. Although initial consideration had been given to including students from Grades 4 and 5 classes in both years of the study, it became obvious that this would complicate further the administrative problems for principals. As a result, the study was conceived for Grade 4 classes the first school year (1974-75) and Grade 5 classes the second year (1975-76).

Provision for School Size. So that the study would not be limited to large schools which could supply approximately 106 students for all four sizes, the design was modified to include smaller schools that could provide 53 students. One year they would have sizes 16 and 37, the other year sizes 23 and 30. An attempt was made to pair schools

with similar characteristics so that the second member of the pair could provide the two complementary sizes in the set of four each year.

Contacting the Schools

It was desirable to have participating schools from each administrative area in the City in order to provide a variety of schools. Area Superintendents arranged to outline the project to their principals and to identify those who might be interested. The researchers then met individually with principals to discuss details of the project with them and to examine any specific problems and unique organizational characteristics of their schools. If they were then able to participate, a meeting with the staff of a school was usually arranged. The project was outlined to the staff, details of its operation and design were then discussed, questions were answered, and particular problems which might apply to a given school were raised by the teachers.

Although the initial reception of the researchers varied from reserved to enthusiastic, the study was well received. The researchers were encouraged by the interest which the teachers displayed in the study, the concern for design which was manifested by a few teachers in every school, and the thoughtfulness with which they considered the possible implications of the project for themselves, the students, the parents and the school.

No effort was made in any school to have teachers volunteer at the conclusion of a meeting, and indeed, an attempt was made to "undersell" the project, making sure that all the possible conditions were identified. For example, careful consideration had to be given to the two year commitment which involved a grade change, not only for participants, but for non-participating staff in the school as well (i.e. in order to accommodate the shift of participating teachers to Grade 5 for the second year, the incumbent Grade 5 teachers also had

to switch grades). Furthermore, the random assignments of students to class sizes would remove from the school flexibility in matching students and teachers and, depending on the size of the school, could create organizational problems such as split-grades to accommodate the assigned class sizes.

In spite of the various ways principals first approached their staffs, it appeared that teachers were given alternatives and that those who felt they should not participate did not have to. When participating teachers responded to the statement: "I volunteered to participate in this study," 29 teachers "strongly agreed" or "agreed," while four teachers "were uncertain," only one "disagreed," and none "strongly disagreed."

All parents of children who would be participating in the study received a letter informing them of the study (see the Technical Supplement). An important condition communicated to parents was that no child would be in a class size of 37 (or 16) for both years. It was also pointed out that the average class sizes in participating schools during the study would be lower than if the school were not involved in the study.

Falling enrolment, staff changes and the necessity for teachers to make a two year commitment made it impossible for many schools to participate. In one school, after a warm and enthusiastic reception by the teachers, they decided that parental reaction would represent an insurmountable barrier. Because of declining enrolment, some principals felt that they would not have enough students to provide the required class sizes for two years. Many schools which could not participate were able to assist the project by providing classrooms for pilot testing of the observational schedules and other instruments.

Not only were the numerous visits and contacts with principals and staff time-consuming, but the procedure of securing participation on the basis of "informed consent" resulted in an anxious four months. The initial meeting with senior officials in Toronto was held early in November. It was not until the beginning of March that the researchers had met with the staff of the first school which shortly afterwards agreed to be in the project.

The procedures which have been described above are a significant feature of the study. Very few research studies have had an opportunity to provide for extensive preliminary discussion at all levels within a school system before their implementation. The implementation procedures also appear to have helped generate serious support for the study from the teachers in participating schools.

The Assignment of Students and Teachers to Class Sizes

The procedures for assigning students and teachers to class sizes were initiated in the Spring each year of the study, when possible in conjunction with the school's regular promotion meetings.

The basic principle underlying the assignments was to ensure that there be some form of random assignment of students and teachers to classes and that some form of control over the assignments be maintained by the researchers. Within this framework there was some flexibility to handle problems or situations specific to individual schools.

In each school, students were assigned to classes on a random basis, stratified by sex and by a rating of academic performance (high, average or low, relative to performance in the grade in the school). Stratifying the assignments in this fashion allowed proportionate representation by sex and by performance in each class size, and was a safeguard to ensure that no class in the study would be composed mainly of either

the best or worst students in a school, or would have students predominantly of one sex.

Another factor accounted for in the assignment scheme was compatibility (or incompatibility) of pupil-pupil or pupil-teacher placement. For example, where previous experience revealed that it was extremely disruptive to have two particular students in the same classroom, procedures were set up to assign them to different classes. In a similar manner, a specific student could be placed with, or away from, a particular teacher. All special assignment cases were specified before the school knew to which class size a particular teacher or pupil would be assigned. There were only a few schools which felt justified in keeping two pupils apart, or together, or where it was important to have a particular student assigned to or away from a given teacher.

Assignments for the First Year. In order to implement the assignment procedures for the first year of the study, talks were held with the principal of each participating school regarding the way they would be done. A detailed listing of the procedures for assigning students and teachers is provided in the Technical Supplement and only the main steps are reviewed here:

(a) preliminary procedures involved compiling lists of students for classes in participating schools, as well as a deck of cards -- one card for each student, one for each teacher;

(b) a meeting was then held at each participating school to obtain further information (e.g., ratings of students' academic performance and the names of students transferring in and out of the school);

(c) the final step was to update the information on the cards and deal the required number of students from each of six categories of student to each class. Each class was thereby stratified by sex and academic performance;

(d) the teachers were then randomly assigned to each class size.

Assignments for the Second Year. The procedures employed in the assignments the second year were similar to those used the previous year except that two additional factors had to be considered. First, to meet the promise made to parents, it was necessary to add a constraint that no pupils who were in the largest (i.e. 37) or the smallest (i.e. 16) class size the first year would be re-assigned to the same size.

The second factor related to the teacher assignments; the main condition to be satisfied was that no teacher would retain the same class size for the second year. However, the procedures developed for the second year were also based on the following considerations:

(i) Teacher Motivation: One important factor influencing a teacher to volunteer for the study was the opportunity to teach one of the two smaller class sizes;

(ii) Maximize Shift in Class Sizes: In order to provide the opportunity of experiencing different sized classes, it seemed advisable to assign teachers to a class size which varied as much as possible from the first year's size;

(iii) Class Size 30: Teachers assigned to class size 30 were most likely to feel that they had not participated in a class size study.

As a result, it was decided that a teacher who taught a class of size 30 or 37 in Grade 4 would not be assigned to these sizes in Grade 5. Similarly, if possible, a teacher would not be given the opportunity to move from size 16 to 23, or vice versa, in the second year. The specific rules used to accomplish these conditions are presented in the Technical Supplement.

The distribution of students and teachers across class sizes in the two years are presented in Tables 5 and 6. It is necessary to point out that in order to run the study for two years and to meet the conditions of changing teachers' and students' class sizes, some pupils had the same teacher for two years.

TABLE 5

STUDENT CLASS SIZE DISTRIBUTION OVER TWO YEARS

Class Size 1975-76	Class Size in 1974-75				TOTAL
	16	23	30	37	
16	2	24	47	54	127
23	28	13	36	90	167
30	37	19	25	85	166
37	32	73	91	1	197
TOTAL	99	129	199	230	657

TABLE 6

TEACHER CLASS SIZE DISTRIBUTION OVER TWO YEARS

Class Size 1975-76	Class Size in 1974-75					TOTAL
	*	16	23	30	37	
*		1	1	1	1	4
16			1	3	5	9
23	3			4	3	10
30	1	4	2			7
37		3	5			8
TOTAL	4	8	9	8	9	38**

* In study for only one year.

** This total includes all teachers who were ever in the study; some teachers who were only in the study for one year were subsequently dropped from the data analysis.

Procedures for Updating Student Assignments. Procedures for updating the student assignments were required:

(a) to ensure that the assigned class sizes (i.e. 16, 23, 30 and 37) did not deviate by more than ±2 for more than a short period of time;

(b) to monitor transfers in and out of the schools during the year; and,

(c) to maintain for each class an exact student membership file.

In September of the first year, each principal was visited to review with him/her personally the assignments of students. The principals were then given a set of guidelines to follow for managing changes, although they were encouraged to contact the Research Department if they had any problems with the arrangement.

The organizational implication of accommodating the class sizes in the study must be mentioned. In some schools it was necessary to create a split-grade class for non-participating pupils. Furthermore, while most classes satisfied the study criteria (see the section on Monitoring Actual Sizes) the constraint that a class size would not deviate by more than ±2 sometimes caused considerable transferring of pupils from one class to another, particularly in those schools with a high mobility rate. Despite the natural reluctance of principals and teachers to transfer pupils from one class to another during the school year, participating schools were very co-operative in following the guidelines.

In this study, no attempt was made to control the "natural" characteristics of the school setting with the exception of class size. In the normal course of events in most classes some students will transfer in and out regardless of the initial organization, and this is true in the study.

The Assignment of Class Sizes to Schools

Proposed Assignments and Modifications. The design initially was developed to involve four classes in "large" schools and two classes in smaller schools each year of the study. This feature ensured that the sample was not limited to schools which could provide approximately 106 students in one grade in order to carry the set of four class sizes. In a smaller school approximately 53 students were required for two classes. Therefore, students and teachers would be either randomly assigned to class sizes 16, 23, 30 and 37 each year of the study, or in the smaller schools the assignments would be to class size 16 and 37 one year, 23 and 30 the other year.

In schools where two classes were provided, an attempt was made to "pair" the schools so that if one carried sizes 16 and 37 the first year, the second member could be assigned sizes 23 and 30, with this pattern switching for the second year. Factors considered in the process of pairing schools included location of the school and census tract data on household income, percentage of owner-occupied dwellings and percentage of population with English as the mother tongue.

Toward the end of the planning phase of the study, it became apparent that the original proposal would have to be modified to enable a few schools to provide three classes or two classes with sizes different from the above combinations. This allowed for the continued participation of schools in which only three teachers had volunteered to participate in the study, the enrolment was declining, or the number of students in the desired grade was greater than 53, but fewer than 106. Modifications based on such factors were required in both years of the

study and resulted in patterns of class size assignments in schools such as those illustrated in Table 7.

TABLE 7

PATTERNS OF CLASS SIZE ASSIGNMENTS

School	1974-75 Class Size				1975-76 Class Size			
	16	23	30	37	16	23	30	37
Original Design								
Example 1 (large school)	X	X	X	X	X	X	X	X
Example 2 (small school)	X	-	-	X	-	X	X	-
Modified Design								
Example 1	-	X	X	X	X	-	X	X
Example 2	-	-	X	X	X	X	-	-

Monitoring the Actual Sizes of Each Class. During each year of the study, the actual sizes of all participating classes were obtained. These data were necessary to determine whether a class satisfied the criteria for its assigned size, and consequently decisions could be made concerning any required reassignments or omissions prior to data analysis.

Although the rule of thumb communicated to participants was that the acceptable variation in class size was ±2, each size could actually vary by 3 before any overlap between two assigned class sizes occurred. The intervals for the assigned class sizes are illustrated in Table 8.

TABLE 8

CLASS SIZE INTERVALS

Assigned Class Size	Interval (±2 Constraint)	Interval (±3 Constraint)	Real Limits* of Interval (±3 Constraint)
16	14 - 18	13 - 19	12.5 - 19.5
23	21 - 25	20 - 26	19.5 - 26.5
30	28 - 32	27 - 33	26.5 - 33.5
37	35 - 39	34 - 40	33.5 - 40.5

* The Real Limits take into account continuous data which might arise when class sizes from successive visits are averaged. For the interval 13 - 19 for example, while it may seem that the smallest number going into the interval is 13 and the largest 19, the numbers 12.6, 12.7 and 19.1, 19.2 would also go in the interval should they occur in the data when the average is computed. The real limits of this interval include values between 12.5 and 19.5.

In order to monitor the actual class sizes, the classroom observers obtained the enrolment figures for each class eight times during each school year (from October to May). The following data were then computed for each participating class: mean number of pupils enrolled, standard deviation, and range (the interval formed by the lowest and the largest size recorded for the year).

A summary of the enrolment data for the four groups of assigned class sizes is presented in Table 9. For each of these four groups of classes, the mean size is within the ±2 interval. However, values of the range indicate that there were some classes with enrolments varying by more than ±3 from its assigned class size.

TABLE 9

SUMMARY OF ENROLMENT DATA BY ASSIGNED CLASS SIZE

Assigned Class Size	First Year (1974-75) Mean Size	Standard Deviation	Range	Second Year (1975-76) Mean Size	Standard Deviation	Range
16	16.63	1.36	12-20	16.76	1.28	14-19
23	23.66	1.44	20-27	23.54	1.89	19-27
30	30.34	2.08	26-35	28.64	2.07	25-33
37	35.74	1.57	33-39	35.80	1.79	32-39

For each of these classes, the complete sequence of enrolment data was examined in detail. While all the data are provided in the Technical Supplement, several of the classes with enrolments varying by more than ±3 will be presented here (see Table 10) to illustrate the main patterns that occurred and the subsequent decisions that were made:

> Example 1: The sequence illustrated in this example occurred a few times for each class size. Basically, in these cases the variation outside the constraint of ±3 occurred once or at most twice. Because most of the enrolments were within the ±3 constraint, these cases were not serious and no changes in assigned class sizes were made.
>
> Example 2: This pattern occurred twice, once for class size 30, in addition to the one illustrated here for class size 37. In both these cases, the mean class size was exactly mid-way in the real interval between its assigned size and the next lowest size. However, since the drop by -4 students occurred in the last half of the school year (i.e. late February), it was decided that no change from the assigned class size would be made.
>
> Example 3: This pattern was the most serious, but it occurred only once. In this case, the mean class size was outside the real limits of size 37 and within the limits of size 30. In addition, the drop by 4 students occurred relatively early in the school year (mid-November) and so this class was reassigned to size 30.

TABLE 10

SEQUENCE OF ACTUAL CLASS SIZES
(EXAMPLES FOR CLASSES WITH ENROLMENTS VARYING BY MORE THAN ±3)

Assigned Class Size[a]	±3 Interval		Mean Class Size	Sequence of Observation Visits #1 (Oct)	#2 (Nov)	#3 (Nov)	#4 (Jan)	#5 (Feb)	#6 (Apr)	#7 (Apr)	#8 (May)
37	34 - 40	eg.1	34.875	33*	36	35	35	35	35	35	35
		eg.2	33.500[b]	34	34	34	34	33*	33*	33*	33*
		eg.3	33.375[c]	35	34	33*	33*	33*	33*	33*	33*[d]

* Indicates a class size in the range outside the ±3.

a For convenience, the examples choosen are for classes assigned to size 37.

b Mean class size is exactly mid-way in the real interval between its assigned class size and the next smaller assigned class size.

c The mean class size is outside real limits of assigned size 37 and within limits of size 30.

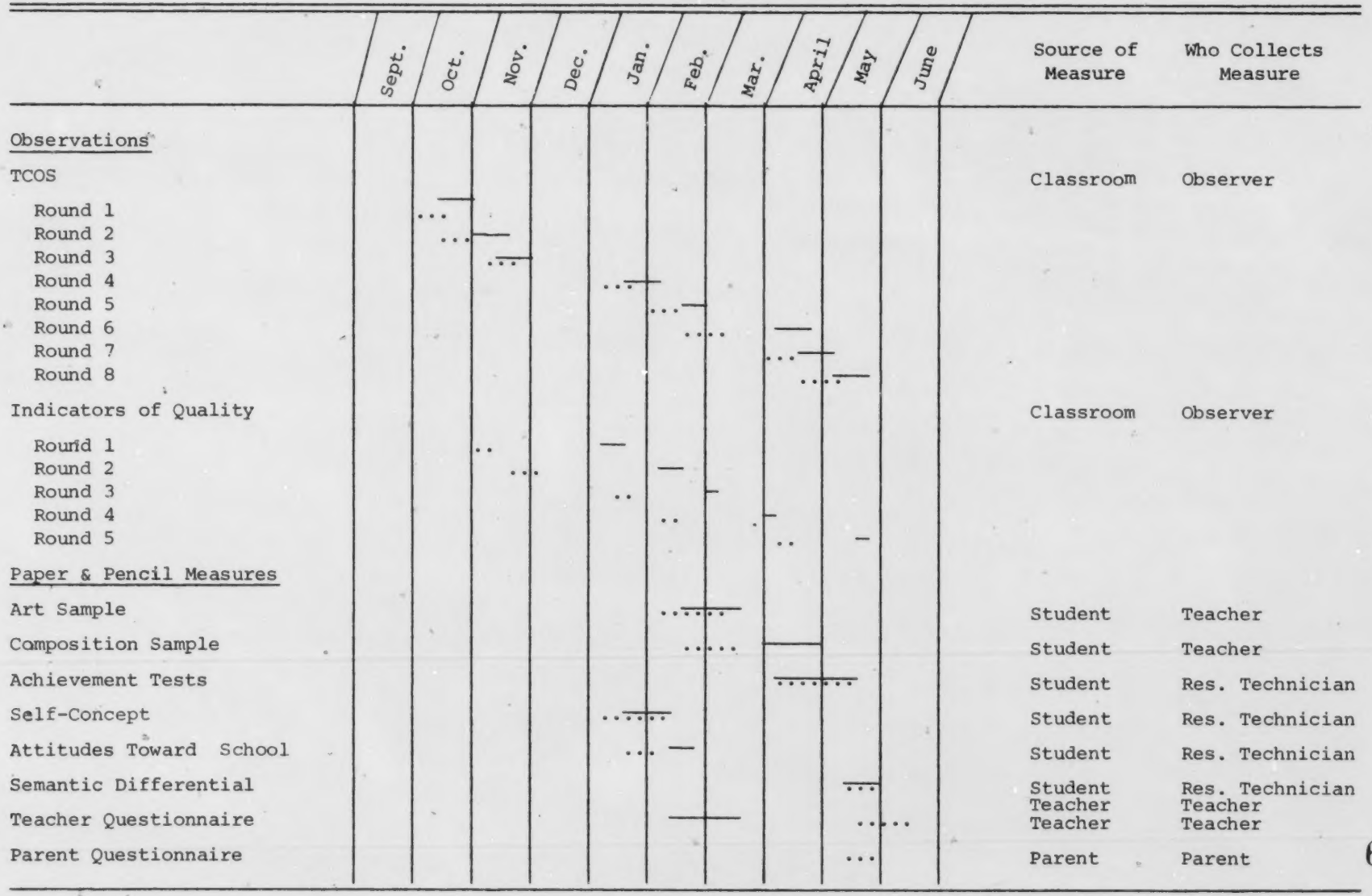
d This sequence necessitated a reassignment of this class to size 30.

Scheduling of Activities

Table 11 summarizes the data collection activities undertaken in the study. Attempts were made to schedule these activities through the school year so as to reduce interference where possible. However, for many measures (e.g., student achievement) it is best to wait until the latter part of the school year to increase the chance of identifying effects of the different class sizes. Activities generally were not scheduled toward the beginning and end of the school year or right before and after major school holidays and the resulting schedule was heaviest from April to mid-May.

TABLE 11

TIMETABLE OF DATA COLLECTION ACTIVITIES IN THE SCHOOLS
(SEPTEMBER, 1974 - JUNE, 1976)



——— First year of the study (Grade 4, 1974-75).

····· Second year of the study (Grade 5, 1975-76).

Scheduling of Observations. For the TCOS, classes were not observed on Monday morning or Friday afternoon; it was decided that, if possible, each class should be observed once on each of the eight remaining half-days in the week. Each year the teachers were notified in advance of the times and dates of the first visits to their classrooms, but no notice was given of the subsequent visits. Every month informaton was obtained from the schools about field trips and other variations from normal class schedules, so that rescheduling of missed visits could be reduced to a minimum. In both years, all observers visited each class at least once and no more than one observer was allowed to make the maximum of three visits to the same class.

Participating classes were observed five times each year with Indicators of Quality. Because all classes in a school were observed on the same day by the same observer, the Indicators of Quality scheduling could not be as rigorous as the TCOS scheduling. These observations were scheduled for three days of the week (Tuesday, Wednesday, and Thursday). In the first year of the study, the observations were delayed until January so that personnel from the Institute of Administrative Research could come to Toronto to provide the training session for observers. The five observation rounds were distributed more evenly during the second year of the study.

Scheduling of Paper and Pencil Measures. The student achievement testing was scheduled as late in the school year as possible, with the constraint that no testing be conducted in June. The achievement testing required two two-hour sessions in each class (one session for vocabulary and reading, the second for the two mathematics subscales). The achievement data were collected by research technicians who received training before administering the tests in the school.

The instruments designed to measure students' self-concept and attitudes and perceptions to school were administered somewhat earlier in the school year than the achievement measures (e.g., the self-concept measures were administered in mid-January). This was done on the assumption that size effects on the latter two measures should be less susceptible to length of exposure to a particular classroom and so that not all the testing would be concentrated toward the end of the school year. Research technicians administered these questionnaires and scales since it was felt that students would feel less restricted and more open in responding than if the classroom teacher were to do this.

On the other hand, it was felt that no visits by the researchers were required for the art and composition samples and so these activities were organized by the teachers according to detailed instructions. The teacher questionnaire which was administered just prior to the March break in the first year was administered later the next year in response to teachers' requests.

Overview of Procedure and Design

The study involved class sizes 16, 23, 30 and 37, ± 2. In the first year of the study, students and teachers were randomly assigned to classes of these sizes in Grade 4. The student assignments were stratified by sex and academic performance. In the second year, the assignments to Grade 5 classes were made in a similar fashion with the constraint that:

(a) students could not be in a class size of 16 or 37 for the two years of the study; and,

(b) teachers be given the opportunity of teaching in two different sized classes.

The design initially was developed to involve four classes in "large" schools (those providing 106 students in one grade) and two

classes in "smaller" schools (providing 53 students) each year of the study. Attempts were made to "pair" the smaller schools so that complementary class sizes could be carried by the second school each year. However, neighbourhood schools with similar characteristics could not be found to participate as a pair for smaller schools in all cases and modifications were made to enable a few schools to provide three classes each year.

The actual size of each participating class was monitored eight times per year to ensure that the size constraints were satisfied. The data analyses are based on a total of 62 classes in nine sets of schools* with class sizes each year, distributed as shown in Table 12.

TABLE 12

NUMBER OF CLASSES BY CLASS SIZE EACH YEAR

	Class Size			
	16	23	30	37
First Year	7	7	8	8
Second Year	9	9	7	7
TOTAL	16	16	15	15

In order to investigate the effects of class size, a variety of measures were taken including: observations of classroom activities, students' opinions and attitudes, self-concept and achievement, teachers' opinions and attitudes, and parental opinions.

* Two of these nine sets involved "paired" schools. In one school in the original sample, one of the two participating classes couldn't satisfy the size constraints early in September of the second year and so this class was dropped from the study (because of lack of flexibility in the non-participating classes). Since only one valid class would have remained, both classes in this school are excluded from the data analysis. In a similar fashion, two teachers who participated for only one year in other schools are excluded. A breakdown of the 62 classes by class size for each participating school is provided in the Technical Supplement.

Results

The results have been organized in five sections which reflect in part a natural sequence. The first section presents the initial expectations of the teachers, as well as their opinions and attitudes after their experiences with different class sizes. The next section deals with the classroom process variables -- data collected by observing the classrooms. The opinions and attitudes of the students are presented next. The fourth section -- student outcome measures -- as well as presenting achievement test data, also contains results from measures of Composition, Art and Student Self-Concept. The final section deals with parents' opinions and attitudes as expressed a few weeks before the completion of the project.

Teachers' Opinions and Attitudes

Teachers' Expectations of the Effects of Class Size

Expectations Prior to the Study. Before the classes were formed teachers were asked (through an open-ended question) to describe the effects they expected in each of the proposed class sizes. The responses were subsequently classified according to tone (positive, negative) and according to content (e.g., expected effects on student outcomes) and are presented in Table 13.

First, the results clearly show that teachers had very positive expectations regarding likely effects of the smaller class sizes (i.e. 16, 23) and negative expectations regarding the larger classes (i.e. 30, 37). The content of the comments was classified according

TABLE 13

NUMBER AND CONTENTS OF THE TEACHERS' COMMENTS REGARDING EXPECTATIONS OF THE EFFECTS OF CLASS SIZE*

	Class Size			
	16	23	30	37
(a) Student Outcomes				
- academic improvement	12	6	0	0
- non-cognitive effects: more involvement, self-confidence	6	6	0	0
(b) Process/Environment and Teacher Effects				
- more teacher-pupil interaction, more individual attention to pupils' problems, better rapport	40	30	0	0
- more pupil-pupil interaction and group solidarity	9	7	1	0
- less formal teaching	2	1	0	0
- more relaxed environment, more materials and space	8	14	2	0
- better discipline	4	2	0	0
- time to make aids, change methods	2	0	0	0
(c) Program Effects				
- more individualized program	15	12	0	0
- more remedial help, enrichment	8	17	0	0
- more extracurricular activities, field trips and experience learning	10	7	0	0
- more small group work possible	3	7	1	0
(d) Other Comments	1	4	8	4
TOTAL	120	113	12	4

	Class Size			
	16	23	30	37
(a) Student Outcomes				
- decreases academically	0	0	3	3
- non-cognitive effects: little involvement, frustration, dependence	1	0	9	9
(b) Process-Environment and Teacher Effects				
- less teacher pupil contact, less individual attention, little time for pupils' problems	0	0	21	28
- class split into social groups; poor socially	1	0	1	1
- more structure, formality and seat work	0	0	8	13
- crowded environment, less space and equipment	0	0	1	3
- discipline difficult, less co-operation from pupils	1	2	1	4
- heavy workload, strenuous work situation	0	0	1	6
(c) Program Effects				
- more group-oriented program	0	0	3	8
- less remedial help, little in depth study	0	0	3	11
- limited extracurriculuar activities, field trips and experience learning				
- less group work possible or larger group required	1	0	7	8
(d) Other Comments	2	8	1	2
TOTAL	6	10	60	97

* Responses of 34 teachers to four open-ended questions regarding possible effects of each class size. Categories were combined so that the number of responses in a category may exceed the number of teachers.

to twelve specific topics dealing with either student outcomes, process/ environment and teacher effects, or program effects. The theme receiving the most emphasis might be termed "individualization." Teachers had high expectations that class size would affect teacher-pupil interactions. They felt that the smaller classes would enable them to give more individual attention to pupils' problems and allow them to develop better rapport with pupils. On the other hand, they felt that the larger classes would result in less individual attention and little time to deal with pupils' problems. They also suggested that there would be more group solidarity in the smaller two sizes.

Still on the theme of individualization, many comments related to the effects of class size on the type of program that could be offered. Many teachers suggested that the smaller classes would allow for the introduction of an individualized program (as opposed to a more group-oriented program, aimed at the "average" pupil in the larger classes) with more remedial help and opportunities for enrichment (as opposed to a less varied program in the large classes with little in-depth study possible) and with more extra-curricular activities, field trips and experience learning (with these activities severely limited in the larger classes).

Two other noteworthy themes of the teachers' comments related to academic improvement and non-academic gains (i.e. more involvement, self-confidence, independence, enjoyment). Pupils in the smaller classes were expected to benefit with more academic improvement and non-academic gains as opposed to pupils in the larger classes. Furthermore, teachers were looking forward to a more relaxed, enjoyable environment in the smaller classes, with more structure, formality and seat work in the larger classes.

Expectations after Experiencing Various Class Sizes. In order to discover whether experience with different class sizes altered their expectations, teachers were again asked the same question toward the end of both the first and second years of the study. In analyzing these responses, the initial expectations (i.e. prior to the study) of an individual teacher were compared to those of the same teacher after experience with a particular class size. For example, if a teacher had size 23 in the first year of the study, that teacher's initial expectations for the smaller classes* were compared to the responses for size 23 on the first year questionnaire, then again on the second year questionnaire. If the teacher had size 30 the next year, the initial expectations for larger classes were compared to the responses for size 30 on the second year questionnaire. In this way, it was possible to compile a list of initial predictions which were still maintained as opinions after experience with a given class size.

The analysis of these responses indicated that teachers' expectations remained quite stable after direct experience with the class sizes. For example, by grouping responses to the question including all four sizes, we find 43 initial statements to the effect that class size would determine the amount of individualization possible. The next two years these teachers, after their experiences, made 35 statements that could be identified as matching or as a re-statement of the original view that class size actually had affected individualization. These statements which were consistent suggest that size 16 and 23 afforded a small enough number of pupils for:

---

* Because teachers tended to be consistent in referring to sizes 16 and 23 as "small classes" and to sizes 30 and 37 as "large classes," the initial responses were divided into two groups for this analysis.

(1) individual attention to academic progress;

(2) individualization of classroom program;

(3) immediate attention to pupil problems;

(4) personal interest in each pupil as another human being.

In contrast the teachers expected individualization to be more difficult in size 30 and almost impossible in size 37. Eight teachers expected that pupils' achievement would be higher in the small classes than in the large classes; after experience with small classes there were seven of them who felt there had been an improvement, explaining that smaller classes gave greater opportunity for academic supervision. Nevertheless, some expectations of the teachers turned out to be less resistant to change after direct experience with a particular class size. Fourteen teachers thought an experience approach to education and opportunities for enrichment, field trips and discussions would be more easily instituted in classrooms of size 16 and 23; only four supporting statements were recorded after experience with smaller classes.

Rapport in the classroom, either teacher-pupil or pupil-pupil, was mentioned fifteen times in response to the preliminary questionnaire and although rapport was mentioned fourteen times, there was less unanimity about the class size preferred. For example, size 16 was initially preferred for its group solidarity. However, after experience some teachers observed that in size 16, a child deviating from the norm on any characteristic was more obvious and more vulnerable to social pressure. On the other hand, it was observed that socially withdrawn children tended to be overlooked in the larger classes.

Examples of four initial expectations where individual teachers later altered their opinions follow:

(1) size 16 would be a close-knit group: the small group became too close-knit to be positive;

(2) size 16 would be good for pupil rapport: size 37 had the best class spirit;

(3) size 16 would put less pressure on pupils: in size 16 there was too much pressure from the peer group and from teacher supervision;

(4) size 37 would have better enrichment opportunities for bright children: size 16 actually turned out to be best for bright children.

Teachers' Opinions after Experience with a Class Size

A series of questions asked for teachers' opinions about the class size they had experienced. The distribution of responses is classified by class size in Table 14. An examination of this Table shows that there are distinct response patterns wherever there is a significant χ^2. Teachers of the smaller classes tended to feel that:

- the quantity of available teaching aids was superior; χ^2 (6) = 17.4, $p < .01$;
- the display area in the classroom was superior; χ^2 (6) = 25.3, $p < .001$;
- the floor space in the classroom was superior; χ^2 (6) = 38.4, $p < .001$;
- the blackboard area was superior; χ^2 (6) = 18.7, $p < .01$;
- fewer students had difficulty hearing; χ^2 (3) = 20.7, $p < .001$;
- fewer students had difficulty seeing; χ^2 (3) = 32.8, $p < .001$.

The pattern of responses did not differ significantly among class sizes for the other variables such as the quality of available teaching aids and satisfaction with the storage space in the classroom.

TABLE 14

FREQUENCY DISTRIBUTIONS OF TEACHERS' OPINIONS ABOUT CLASS SIZE THEY EXPERIENCED

Question & Response Category	Class Size 16	23	30	37	χ^2
The quantity of available teaching aids is:					
superior	10	8	1	1	
adequate	6	5	10	7	
inadequate	0	2	3	7	17.42*
The quality of available teaching aids is:					
excellent	10	6	3	2	
adequate	6	5	9	9	
inadequate	0	4	2	4	9.35
The display area in the classroom is:					
superior	8	10	2	0	
adequate	7	4	10	4	
inadequate	1	1	2	10	25.29**
The storage space in the classroom is:					
superior	7	7	2	0	
adequate	5	6	5	6	
inadequate	4	2	7	9	11.09
The floor space in the classroom is:					
superior	16	9	1	0	
adequate	0	4	4	3	
inadequate	0	2	9	12	38.39**
The blackboard area in the classroom is:					
superior	16	10	3	4	
adequate	0	3	7	7	
inadequate	0	2	4	4	18.68*
Do students have difficulty hearing because of size of class?					
number with difficulty	4	2	23	30	
number without difficulty	252	343	397	525	20.7**
(Proportion with difficulty)	(.016)	(.006)	(.058)	(.057)	
Do students have difficulty seeing because of class size?					
number with difficulty seeing	9	21	34	80	
number without difficulty seeing	247	324	386	475	32.77**
(Proportion with difficulty)	(.036)	(.065)	(.088)	(.168)	

* $p < .01$
** $p < .001$

Teachers were asked whether they had made any changes which were directly related to the size of their class on a list of areas which included class management, supervision, use of audio-visual aids, use of print media, handling supplies, physical layout of the class and evaluation. The percentage of teachers reporting that they made changes is presented by class size (Table 15).

First, examining the totals by class size, the largest percentages of teachers to indicate that they had made changes were those who experienced size 16 (77 per cent) and size 37 (65 per cent) with fewer changes being reported by teachers of size 23 (53 per cent) and the least number of changes being reported by teachers of size 30 (37 per cent).

Second, examining the data by area of change, the largest percentages of teachers indicated that they had made changes in physical layout (82 per cent), evaluation (72 per cent), class management (71 per cent) and supervision (66 per cent). Comments made by teachers regarding these areas are summarized below:

> Physical Layout. Teachers of size 16 commented that they were pleased with the amount of room available, especially for centres and activities and they reported that pupils had a choice of where to work and were free to move around; teachers of size 30 reported that grouping was beginning to get tight; and for size 37, teachers reported that the rooms were crowded and desks took up too much room.
>
> Evaluation. Teachers of sizes 16 and 23 were happy because marking took little time, the advantage being that corrections were immediate. Some alternative methods were mentioned such as interviews and pupil self-evaluation and it was felt that evaluation could be more personal and involved, more extensive.
>
> In size 30, it was reported that marking becomes more formal, time-consuming and is sometimes delayed. In size 37, marking and evaluation were reported to be overwhelming. Teachers enlisted pupils for marking (either their own work or others'). Evaluation was based on tests or specific projects, rather than on general teaching knowledge of pupils. Teachers seemed dissatisfied with this kind of evaluation.

TABLE 15

PERCENTAGE OF TEACHERS REPORTING CHANGES DIRECTLY RELATED TO CLASS SIZE
(ACTUAL NUMBERS REPORTED IN PARENTHESES*)

Area of Change	Class Size 16	23	30	37	Total (by Area of Change)
(a) Class management	94% (15/16)	60% (9/15)	46% (6/13)	83% (10/12)	71% (40/56)
(b) Supervision	94% (15/16)	47% (7/15)	38% (5/13)	83% (10/12)	66% (37/56)
(c) Use of audio-visual aids	40% (6/15)	40% (6/15)	23% (3/13)	23% (3/13)	32% (18/56)
(d) Use of print media	53% (8/15)	40% (6/15)	25% (3/12)	31% (4/13)	38% (21/55)
(e) Handling supplies	67% (10/15)	31% (4/13)	23% (3/13)	62% (8/13)	46% (25/54)
(f) Physical layout	100% (16/16)	87% (13/15)	46% (6/13)	92% (12/13)	82% (47/57)
(g) Evaluation (Marking	81% (13/16)	67% (10/15)	62% (8/13)	83% (10/13)	72% (41/57)
(Other	(5/6)	(4/8)	(2/7)	(5/6)	
TOTAL	77% (88/115)	53% (59/111)	37% (36/97)	65% (62/95)	

* Not every teacher responded for each topic so that the denominators are not always the same.

Management. Teachers of size 16 were happier because they were able to run their classrooms more smoothly and efficiently, with little effort on their part. Children were more responsible in all areas, including working on their own, handling supplies and behaving appropriately.

In size 37, it was reported that rules were set up and had to be enforced. Management was more formal with restricted movement.

Supervision. In size 16 it was reported that there was little need for the teacher to supervise; often the children made supervisory decisions. Although only one teacher actually mentioned being more tolerant of the children's actions, the feeling of more tolerance was evident in the majority of replies. The comments for size 37 suggested that discipline was a problem for the teachers who had to police behaviour, rather than change it. The classes were more structured, the teachers strict. There was no feeling of freedom for the teacher.

As indicated in Table 15, low percentages of teachers reported that they had made changes in the use of audio-visual aids, use of print media and the handling of supplies.

Teachers' Opinions after Experiencing a Class Size Change

Several questions asked teachers for their opinions relative to what they felt or did the previous year (e.g., "I gave more attention to individual pupils this year than last year"). It was decided to analyze these questions according to two "size change" groups, with the first group consisting of those teachers (n = 15) who went from a large class size (30, 37) the first year to a small size (16, 23) the second year; the second group contained those teachers (n = 13) who had a large class size (30, 37) after previously experiencing a small class size (16, 23).

The distribution of responses of these two groups, presented in Table 16, was extremely clear-cut. In comparison to the group changing from small to large class sizes, teachers who went from a

TABLE 16

TEACHERS' OPINIONS ACCORDING TO THE SIZE CHANGE THEY EXPERIENCED IN THE SECOND YEAR OF THE STUDY

Question and Response Category	Size Change Group: Large to Small	Size Change Group: Small to Large	χ^2
Compared with last year, I like this class:			
more	13	4	35.2*
less	2	9	
My energy level compared to last year is: [a]			
higher	8	1	76.4*
same	5	3	
lower	1	9	
I gave more attention to individual pupils this year:			
agree/strongly agree	14	0	167.2*
disagree/strongly disagree	1	13	
My pupils this year contribute:			
more	14	1	155.1*
equally	1	4	
less	0	8	
My pupils this year pay attention:			
more	13	1	91.0*
equally	1	3	
less	1	9	
My pupils this year are satisfied:			
more	12	1	111.8*
equally	3	4	
less	0	8	

* $p < .001$

a One teacher did not respond to this question.

large class size the first year to a small size the second year were significantly more likely to:

- like this year's class more than last year's χ^2 (1) = 35.2, $p < .001$;
- report a higher energy level this year; χ^2 (2) = 76.4, $p < .001$;
- report that they gave more attention to individual pupils this year; χ^2 (1) = 167.2, $p < .001$;
- feel that their pupils contributed more, paid more attention and were more satisfied this year; χ^2 (2) > 90 in each case, $p < .001$.

Teachers' Comments on Various Issues

Teachers were asked to comment on various issues, drawing on their experience with a particular class size. A list of issues was compiled, based on the themes mentioned by the teachers when they replied to the preliminary questionnaire before the classes were formed. The list of issues is rank ordered according to the response rate to each item (Table 17).

The issues which were most often discussed were teacher-pupil relationships (62 per cent), discipline (57 per cent), followed by opportunities to provide an individualized program and methods of teaching (43 per cent each). The lowest response rates were to the issues of skill development (15 per cent), social and learning environment (20 per cent each) and pupil involvement in planning and amount of group work (21 per cent each).

For the four issues receiving the largest response rate, the following sample of teachers' comments were selected as representative of the feelings of the majority in each class size group:

Teacher-Pupil Relationships.

Teachers considered a warm, close relationship with each child desirable.

Size 16. The teacher gets to know the child. This is a beautiful experience. The teacher learns how not to pressure the child, but how to make him feel secure and wanted. The teacher uses the good rapport to stimulate learning. Co-operation is much greater and better.

Size 23. Although I do not feel any more affection or closeness to this group as a whole, I find that I do really know all my kids. When I look back, I suppose I only ever really knew 20 or so of my kids in the large classes I have always had.

It is such an advantage. You know their problems, likes, dislikes and can more easily reform attitudes.

The children I know all feel very close to me and all will come to me for help or to talk or to confide or just be near.

Size 30. With a class of 30, I am not pleased with the amount of time I have to spend with each child. I find it very frustrating not being able to get around and talk to the children individually as much as I would like.

Size 37. In the class of over 28 it has been my experience that you do not get to know any child well. Those who are shy you rarely talk with. Your main concern is marking, planning and discipline.

The pupils themselves become blank faces you are talking at, receptacles you are cramming knowledge, skills, morals into.

The few children you do get to know and achieve an understanding with are then considered teachers' pets.

Discipline

Teachers preferred the more easy-going, democractic classroom possible in small classes and did not enjoy being the more arbitrary ruler often necessary in large classes.

Size 16. Discipline in the small class (16) has been a matter of absolute simplicity. Because of the small size, any devious forms of activity have become virtually impossible. I have maintained the same standards as in the previous year and have been able to exceed them with much less effort and a more relaxed and productive classroom has been the end result. The children also are more aware of the fact that any disruption affects the entire class. The children are generally more responsible in their actions and the smaller numbers mean that there is more material available for them when they are finished their assigned work; whereas, in a larger class with limited materials many are left with nothing to do between the time they finish their work and the time that I can attend to them. The smaller class has the advantage that there are generally a smaller number of troublemakers in the class, plus the fact that should a large number of problems turn up, it is easier to deal with them in a manner that is fair to the entire class.

Size 23. The children participated in their own disciplinary activities better in the smaller class. There is more time in a small class to make sure each child receives equal attention. When a fairly even amount of attention is distributed to each child, the group as a whole seems to have better discipline. There is a stronger pride in accomplishment and organization. The benefits to the group as a whole make each child more aware of his individual contribution.

Size 30. I tended to be stricter in regards to following classroom rules and routines. There seemed to be more "policing" in the larger class. The rapport I had with each pupil after a few months (3) tended to help me control them and influence them.

Size 37. There had to be more rules and set methods of carrying out different procedures because of the larger number of children. They were a very good group of children however, and I could reason with them and we could discuss the need for a certain pattern of behaviour. There were always some who did not co-operate and the larger number made lines in the halls or lines on the sidewalk (on field trips) much more difficult to supervise. The noise level was much higher than with the smaller group and thirty-six children and myself were a great many people in one classroom.

Methods of Teaching

Teachers felt that a small class size by itself did not promote a more individualized method of teaching. Teaching, especially in class sizes of 23 up, was mainly to groups of children.

Size 16. I have had both a totally individualized program and a schedule involving five small groups in reading. Mathematics I have taught as a whole group, mainly because it is the one subject I find most uninteresting and difficult to teach. This method worked well for me as I was then free to assist individually other problem areas. The rest of the day was spent on individual work chosen by the students, prompted by myself, to explore in greater depth. The main focus of their endeavours always seemed to return to creative writing and drama as a means of self-exploration and expression.

Size 23. (Very few comments, none could be selected as representative.)

Size 30. With the larger class I tended to teach it as a whole unit in all areas except reading and spelling. For reading I divided the class into four groups and for spelling I divided the class into two groups. Later in the year I returned to teaching and spelling to the whole class. I always have informal temporary grouping on proof-reading, creative writing, science or social studies, but with the larger numbers I tended not to do this as often. This year, without the few number I tended to take less excursions.

Size 37. Group teaching is the only method suitable to a large class. This kind of group instruction and the pressures to accomplish a certain program during the year necessitate teaching to the average child and helping the slower ones. The poor bright kids are left to their own devices and as a result miss the stimulation they deserve.

Opportunities to Provide an Individualized Program of Instruction

Teachers seemed to have varying concepts of individualization. (See the summary at the end of this section.) Because of this, the quotations on individualization represent a middle-of-the-road view between extremes of individualized programming.

Size 16. There are many opportunities for individual instruction. In fact, most interaction between pupils and teacher is attempted on this basis.

Size 23. (NOTE: Teachers of size 23 felt the opportunity was good for individualization. However, some did not necessarily know how to attempt it with this number.)

In the class size of 25 we had more opportunities to plan and discuss individual project and research work especially with the advanced students. The children were able to work more readily at their own speed. We had more time to discuss and evaluate project work. There was time to discuss the use of research material found in the new library, public libary and encyclopedia and resource books in the classroom.

Greater opportunities. Unfortunately have not taken full advantage of this opportunity due to uncertainty about my own ability to ensure its success. However, greater individual attention has prevented the development of problems in math.

Size 30. I find this very limited by class size. I would like to do more individualized work in the initial teaching situation, but I find most individualized work is done in helping the children while they are doing seatwork.

Size 37. There are few opportunities for individual instruction which is the best and most enjoyable method of teaching and being taught.

I tended to work the class in groups. It was too difficult to set up an individualized program with the number of pupils as well as the many different levels of pupils in the class.

TABLE 17

TEACHERS' RESPONSE RATE TO A LIST OF ISSUES

Rank	List of Issues	% Response Rate
1	Teacher-pupil relationships	62% (38/61)
2	Discipline	57% (35/61)
3.5	Methods of teaching	43% (26/61)
3.5	Individualized program	43% (26/61)
6	Enrichment possibilities	36% (22/61)
6	Remedial work	36% (22/61)
6	Strength of peer ties	36% (22/61)
8.5	Achievement level	30% (18/61
8.5	Structured program	30% (18/61)
10	Amount of informal teaching	26% (16/61)
11.5	Pupil involvement in planning	21% (13/61)
11.5	Amount of group work	21% (13/61)
13.5	Social environment	20% (12/61)
13.5	Learning environment	20% (12/61)
15	Skill development	15% (9/61)

Individualization -- What Is It? Because the teachers in the study seemed to have different concepts of individualization, an attempt was made to define those concepts as found in their comments about the list of issues which was provided (see Table 17). These various "definitions" follow:

(1) A separate academic program for each child in the class based on the child's own choice of topics, his academic level and his own goals.

(2) A separate academic program for each child based on teacher goals for that child.

(3) A separate project or research in some subjects.

(4) Separate attention to sub-groups, such as high or low achievers.

(5) Individual teacher-pupil contact with attention to pupil interest or problems.

(6) Helping individuals in group seatwork situations.

(7) Attention to individual work: "I reviewed individual workbooks more often."

(8) Negative definitions: "A non-group," "non-structure" or "not a traditional" situation, "little formal group work."

Advantages and Disadvantages for High and Low Achievers. Aside from soliciting teachers' comments on the issues discussed previously, teachers were asked to describe the advantages and disadvantages of their particular class size for high and low achievers.

Briefly, summarizing the responses:

Teachers preferred the small classes, although they disliked the social consequences of a small peer group and the isolation of low achievers in a small class.

The teachers preferred to work individually -- "the best method for high and low achievers" -- and the small classes gave them the time to do so. They suggested that the larger classes made individual attention very difficult to the detriment of both high and low achievers.

More details are provided below according to class size.

Advantages for the High Achiever

Size 16. The advantages mentioned most frequently were individual attention (which included immediate marking and correcting, encouragement to achieve goals, a better social and intellectual rapport with the teacher) and the opportunity to progress at an advanced speed (i.e. to achieve "more" academically, to have more special interest projects, to set individual goals and to help plan their program).

Some teachers felt leadership and independence growth was better for high achievers in the small class, but there was dissension on this point.

Size 23. The advantages were mainly attributable to the teacher having time and energy to work with individuals. High achievers could help plan programs at their own level, they could work ahead, have extra interest projects and the teacher would be there to provide challenge at the appropriate time. They did not have the frustration of waiting for others to catch up. Evaluation and corrections were immediate.

The social interaction opportunities were good for the higher achiever because groups could be smaller, there could be more discussion groups and there was opportunity to develop leadership.

Size 30. The advantages were limited. A few teachers felt size 30 was still small enough for high achievers to carry on with independent interest projects and activities. One teacher felt there were "no effects" because bright students would cover the material no matter what the class size.

Size 37. The advantages listed here came mostly from one teacher who felt the high achiever had opportunities for group participation, leadership development, positive competition and social interaction in a large group.

Two teachers felt leadership was forced on high achievers because the teachers were busy with other students, but that this contributed to personality development. One teacher felt size 37 was "not bad" for high achievers because they would be successful anyway. Two teachers said there were no advantages for high achievers.

Disadvantages for the High Achiever

Size 16. The disadvantages were based mostly on a lack of other pupil interaction. High achievers tended to work on their own apart from the group, had less motivation because of a lack of similar pupils and also less stimulating competition.

Size 23. The disadvantages were varied; the small peer group and loss of competition were mentioned, as was the suggestion that a high achiever may be overwhelmed by the greater number of lower achievers (one high achiever tried to do as little work as the lowest achievers in his room). Three teachers agreed that there were no disadvantages.

Size 30. The disadvantages were stated in terms of teacher time. Individual attention, especially for high achievers, was reduced to an unsatisfactory level, even to the point that high achievers were being overlooked. These children, therefore, had to take a lot of responsibility and work on their own.

The academic program tended to be aimed toward the average student. This resulted in less motivation, stimulation, enrichment and challenge for the high achiever, as well as unsatisfactory teacher-pupil rapport.

Size 37. The disadvantages again were based on a definite lack of teacher time. The program became standardized and aimed for the average pupil. There were few, if any, extras. High achievers received little challenge, motivation, or stimulation, and there was not time for meaningful follow-ups or discussions. Behaviour and emotional problems were more pronounced. High achievers were left on their own too much and tended, in some classes, to be bored.

Advantages for the Low Achiever

Size 16. The advantages mentioned most often were the opportunity for individual attention, immediate remedial help and the elimination of frustration because progress could be at the child's own speed. The teacher can be there with encouragement and can catch behaviour problems immediately. The low achiever cannot find a hiding place and therefore has greater supervision.

The low achiever in a small class has a greater opportunity for self-expression and this plus the close attention from the teacher improves his self-esteem and self-confidence.

Size 23. The advantages were much the same as for class size 16 -- the teacher had time for individual attention, encouragement, remedial help and behaviour modification.

Other advantages included more variety in review work, inclusion of the low achiever in small groups and less stigma because of greater class cohesion. One student worked extra hard in the small group to improve so that he "wouldn't be noticed."

Size 30. The advantages included a little individualization, despite the teacher's busy schedule, for remedial help.

Low achievers were more able to hide their academic level in the crowd. However, this size provided more academic peers and an impetus for stimulation. More helpers were available.

Size 37. The advantages were basically nil in most teachers' opinions. One teacher did feel the large size was better for interpersonal relationships and providing motivation and help among the students.

Disadvantages for the Low Achiever

Size 16. The disadvantages related to social problems. The child's low academic status is readily apparent and he becomes self-conscious of his low achievement. His classmates as a group may exert pressure that is too strong. There are fewer children to give him academic help.

It appears that the low achiever's self-esteem in a small class is tenuous and depends on more than just class size for it to grow.

Size 23. The disadvantages mentioned were few. One teacher mentioned that the small number of pupils made the low achiever stand out. Two teachers mentioned "nil" disadvantages.

Size 30. The disadvantages were much the same as for the high achiever -- not enough teacher time for individuals was mentioned most often. Some low achievers were overlooked and amount of remedial work was unsatisfactory. Teaching was aimed at the average student and was in itself reduced by the marking and planning load.

Low achievers did not have much chance for oral participation.

Teachers felt they did not know the students well enough to uncover specific weaknesses readily.

Size 37. The disadvantages were especially emphasized for this group of students because individual attention is essential for their success. The teachers did not have time in class for remedial work and extra time had to be given outside of class or forgotten. Also, individualized programs were rare -- the low achievers had the frustration of trying to keep up with the average. They become frustrated, neglected, fall behind, often doesn't finish. They are easily overlooked and, indeed, can readily hide in the crowd. Overcrowding and its distractions have a detrimental effect. The teacher is not always available at the right time.

The child must cope with larger groups, has little personal time with the teacher and tends to become a behaviour or emotional problem.

Teachers' Responses to the Semantic Differential

Teachers' attitudes toward two concepts "My Classroom" and "The Pupils I Teach" were assessed by creating scales from data collected using the format of the semantic differential technique (see Instrumentation Section, p. 34). Teachers rated "My Classroom" and "The Pupils I Teach" using 18 and 21 pairs of polar adjectives respectively; the adjectives were rated on a seven point scale, with the positive end being scored seven.

The mean positive scale scores for both concepts are presented by class size (Table 18) and they are plotted along with their 95 per cent confidence intervals in Figure 1.* The results revealed a significant overall effect due to class size for the concept "My Classroom," $F(3,58) = 4.2$, $p < .01$. No significant difference was found for "The Pupils I Teach," $F(3,55) = 0.48$, $p > .05$.

TABLE 18

RESULTS OF TEACHERS' SEMANTIC DIFFERENTIALS

Concept	Class Size				Anova	
	16	23	30	37	df	F
My Classroom						
mean	100.9	96.8	92.2	91.4	3,58	4.2*
standard deviation	(6.2)	(5.9)	(10.9)	(9.0)		
Pupils I Teach						
mean	114.1	108.9	110.8	110.2	3,55	0.48
standard deviation	(13.8)	(11.3)	(10.5)	(13.9)		

* $p < .01$.

* A confidence interval provides an estimated range of values which, considering all possible samples, has a high probability (in this case a probability of .95) of covering the true population value.

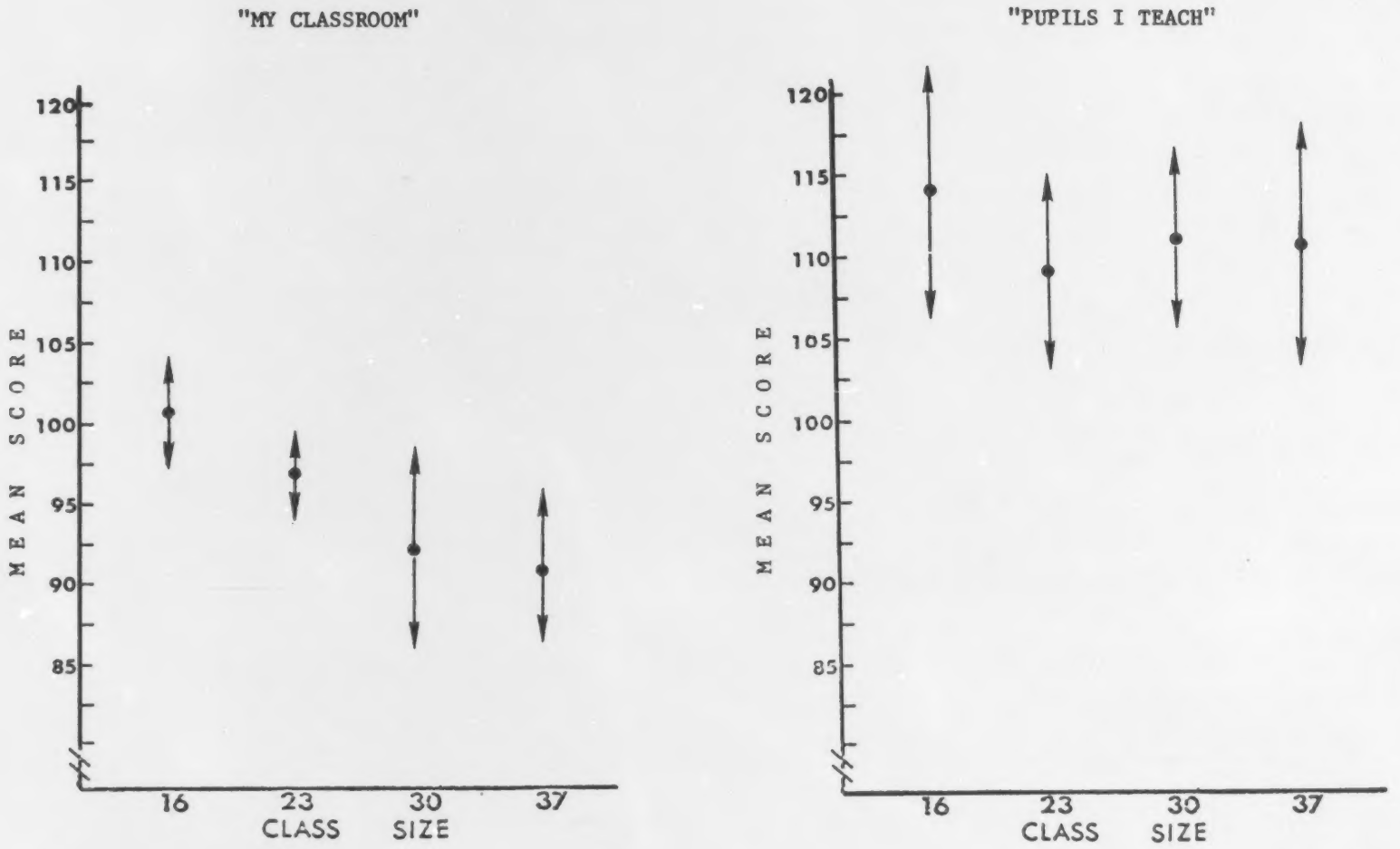


Figure 1. Semantic Differential, Teachers' Forms: Means and Confidence Intervals (95 per cent) by Class Size.

Non-arthogonals pair-wise contrasts were conducted to investigate further the significant effect: teachers who had experienced class size 16 rated "My Classroom" more positively than did teachers from class size 30 ($t54 = 2.88$, $p < .01$) and teachers from class size 37 ($t54 = 3.10$, $p < .01$); both differences were significant.

Observation of Classroom Process Variables

Procedures for the Data Analysis

For the observational data the unit of analysis was the class; that is, mean scores or total scores from each class were compared, rather than individual scores from each observation visit.

Data on the reliability of the observational measures are presented and discussed in the Technical Supplement. With one exception, (the frequency of approving utterances about course content) variables from the timed observations -- Teacher Verbal Behaviour, Individual Pupil Participation and Pupil Participation -- were reliable. Classroom atmosphere ratings were also reliable. However, some of the variables from the remaining sections of the TCOS -- use of educational resources, method of instruction, subject emphasis, physical conditions and conflict resolution -- did not satisfy the criteria for reliability; this is not surprising as most of these variables had low maximum values. Indeed, most of them had a maximum value of eight, as they were counts of the number of observational visits during which a certain event was observed. At any rate, most variables were reliable and some variables in each area of effect were reliable. Although a variable without proven reliability is one of which only imprecise measurements can be

made, these variables were analyzed so that readers could inspect the results themselves. Variables with questionable reliability will be noted in the results and discussed where necessary.

Before the data were analyzed, proportions tests were used to compare the contextual descriptors from each year of the study to see if there were any differences between years in: subject taught; teachers' activity; organization of the class or pupils' task. Cohen's (1969) conventions were used to evaluate the sizes of differences between years. Few differences were observed and these differences were very small; therefore, it seemed that a special selection of data to ensure comparability of years was not necessary. Means of each variable in each year were also compared with proportions tests to see if there were differences between years. Few differences were observed and all but one were extremely small. These data are presented in the Technical Supplement. Scores on the variable for which there was a large difference between years were adjusted; the adjustment is discussed in the following presentation of results.

Differences between class sizes were first assessed by one-way analyses of variance of scores in the four class sizes. When a significant difference was found, pseudovalues were calculated by Mosteller and Tukey's jackknife procedure (1968) and a similar analysis of variance was performed on the pseudovalues*. Class size means were compared by a t-test whose critical value had been raised so that one spurious difference would be expected in the analyses of twenty variables.

---

* The batching of data involved in this procedure helps to balance the effects of schools with different numbers of classes.

One-way analyses of variance were also performed to assess effects due to differences in school. Multiple linear regressions were performed on variables with significant school effects to remove these effects, and the residual scores were then analyzed by size. However, no additional class size differences were detected by this method.

(a) The Toronto Classroom Observation Schedule (TCOS)

Variables in which significant differences between class sizes were detected are listed in Table 19, along with F values and class size means and standard deviation. Similar results of the analyses of variables which were not affected by class size are presented in the Technical Supplement.

Teacher-Pupil Interaction. Probably the most important section of the TCOS in the assessment of teacher-pupil interaction, is the section which categorizes Teacher Verbal Behaviour (TVB).

The following TVB variables, all of which are means with a maximum score of 32, were analyzed.

(1) the frequency of utterances about classroom routines;

(2) the frequency of lecturing ("tell") to all audiences (individual, group or class);

(3) the frequency of lecturing to the class;

(4) the frequency of questioning;

(5) the frequency of approving utterances about course content;

(6) the frequency of disapproving utterances about course content;

(7) the frequency of all utterances about course content;

(8) the frequency of approving utterances about pupils' behaviour;

(9) the frequency of disapproving utterances about pupils' behaviour;

(10) the frequency of neutral utterances about pupils' behaviour;

(11) the frequency of all utterances about pupils' behaviour;

(12) the frequency of all approving utterances;

(13) the frequency of all disapproving utterances;

(14) the frequency of all neutral utterances;

(15) the frequency of all utterances about social matters;

(16) the frequency of all utterances addressed to individual pupils;

(17) the frequency of all utterances addressed to groups*;

(18) the frequency of all utterances addressed to the class;

(19) the frequency of pupil talking (that is, talking which is part of learning);

(20) the frequency of observations during which neither the teacher nor any pupils were talking;

(21) the frequency of all observations during which the teacher was not talking (that is, the sum of variables 19 and 20).

All variables except (5) were reliable.

None of these variables was affected by class size. Of particular interest is the failure of class size to affect the lecturing and audience variables, the frequency of utterances about course content or the frequency of utterances about routines (variables 1, 2, 3, 7, 16, 17 and 18).

One aspect of teacher-pupil interaction is the frequency with which the teacher addresses an individual pupil. Data on the

---

* This variable includes all utterances addressed to any group of two or more pupils which did not contain the whole class. It is not a measure of the frequency of utterances addressed to formal learning groups (reading groups, for example).

interaction of the teachers with individual pupils are collected during the observation of the participation of individual pupils. The score is the proportion of pupils observed who were addressed as an individual by the teacher. It was reliable. Class size did affect this variable. Pupils were more likely to be addressed individually in classes of 16 than in the other three sizes of class, while pupils in classes of 23 were more likely to be addressed individually than pupils in classes of 30 and 37 (there was no difference between classes of 30 and of 37). The linear trend accounted for 96 per cent of the variance between cells. See Table 19 and Figure 2 for the results of this analysis.

Pupil Participation. Data on pupil participation are collected in the sections of the TCOS which categorize Individual Pupil Participation (IPP) and general Pupil Participation (PP). All variables were reliable.

The following IPP variables, all means with a maximum score of 16, were assessed:

(1) the frequency of observations during which the pupil observed was participating (that is, performing the task assigned to him);

(2) the frequency of observations during which the pupil was participating non-verbally;

(3) the frequency of observations during which the pupil was participating verbally;

(4) the frequency of observations during which the pupil was not participating;

(5) the frequency of observations during which the pupil had no task to perform (this variable includes lining up to have work checked).

The PP section provided an estimate of the proportion of students in the classroom who were participating. None of these six measures was affected by class size.

TABLE 19

SIGNIFICANT RESULTS OF THE ANALYSIS OF VARIANCE OF OBSERVATIONAL VARIABLES

Area of Effect	Variable		Class Size 16	23	30	37	F*	Jackknife F**
Teacher-pupil interaction	Proportion of pupils addressed as individuals	- mean	.43	.34	.30	.25	14.43	12.96
		- standard deviation	.08	.09	.05	.08		
Method of instruction	Lecture *** (maximum = 8)	- mean	4.19	4.69	3.13	3.80	3.63	4.06
		- standard deviation	1.11	1.30	1.13	1.78		
Method of instruction	Supervision by teacher while pupils are working*** (maximum = 8)	- mean	1.94	2.44	3.60	2.53	2.94	3.38
		- standard deviation	1.39	1.55	1.50	1.92		
Subject emphasis	Reading (maximum = 8)	- mean	4.38	3.69	2.60	3.13	2.95	2.82
		- standard deviation	1.82	1.70	1.59	1.85		
Use of educational resources	Number of written aids used*** (maximum = 40)	- mean	15.06	17.81	17.93	15.27	4.71	4.53
		- standard deviation	2.24	2.40	2.99	3.61		
Use of educational resources	Proportion of available written aids used***	- mean	.40	.48	.48	.41	4.61	4.50
		- standard deviation	.06	.06	.09	.09		

* At the .05 level the critical F ratio is 2.77.
** At the .05 level the critical F ratio is 2.90.
*** The reliability of this variable was not demonstrated.

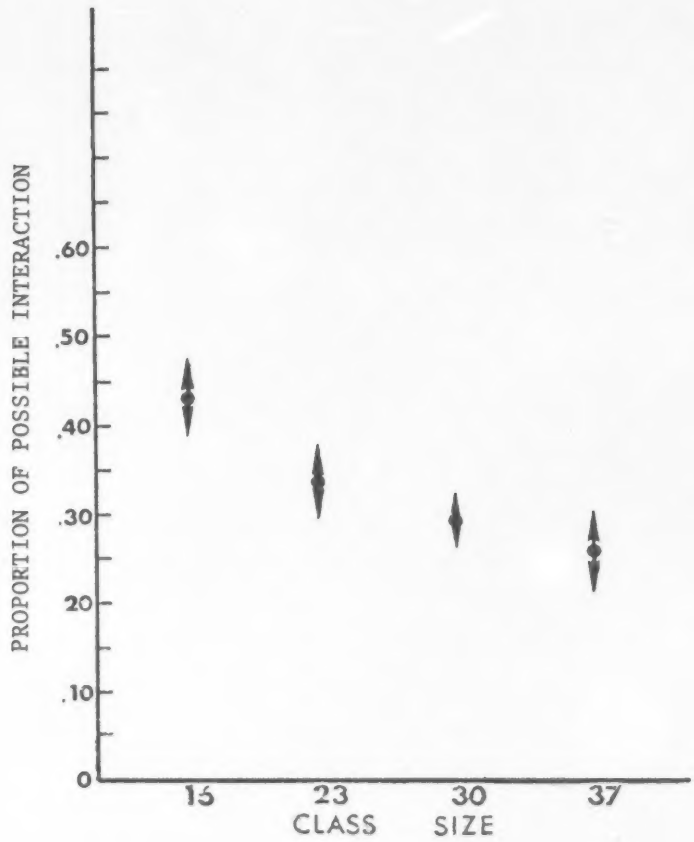


Figure 2. Interactions between the teacher and individual students. (Means and 95 per cent confidence intervals.)

Pupil Satisfaction. It was originally intended that pupil satisfaction should be measured by the section categorizing Pupil Aggressive Behaviour (PAB). However, overt pupil hostility proved to be very rare -- only 66 hostile acts were reported during 1,390 PAB observations, or one every hour and 45 minutes of obervation time. This frequency is so low -- an average of .05 acts per observation -- that any differences observed would be unimportant. However, a similar variable was calculated from the conflict resolution section of the TCOS. The number of observations each year during which no conflict between pupils was reported was calculated for each class. Although this variable was unreliable, it was analysed and no significant differences between class sizes were observed.

Method of Instruction. During the development of the TCOS, the concept of method of instruction was broadened to include non-academic activities, so that "teacher's activity" would probably be a better name for the area of effect investigated. In one section of the checklist the observer noted the three most frequent teacher's activities during the half-days of observation. Counts were made of the number of times the following activities were mentioned in the course of a year: lecturing, organization of learning activities, question and answer, supervision, supervision combined with administrative work (such as marking) and the sum of the two supervision categories.

The organization, question and answer and supervision-with-work variables were reliable, and no differences between class sizes were observed. The reliability of the other variables was questionable, but a difference was found in the lecturing variable. Lecturing was

recorded less frequently in classes of 30 than in classes of other sizes. The linear trend was not significant. See Table 19 and Figure 3 for the results of this analysis. Because the variable's reliability could not be demonstrated, this finding must be interpreted cautiously.

Another section of the TCOS looks at the use of the teacher's time while the pupils are working on their own. The teacher can be classified as working alone, supervising, working with an individual pupil, or working with a group of pupils (a category was included for other activities, but none appeared frequently enough to be of interest). The variables were assessed by counting the number of observation visits each year during which each of these categories was reported, so that the maximum score for a class on each variable is eight. The counts of working alone and working with an individual pupil were reliable and not affected by class size. The reliability of the other two categories was questionable. Working with a group was not affected by class size, but a difference in supervision was found between class sizes. Supervision was more frequent in classes of 30 than in classes of 16; no other differences between class sizes were significant. The linear trend was not significant. See Table 19 and Figure 4 for results. As the reliability of this variable was questionable, it is not easy to attribute this difference to class size. It should be pointed out that if the difference in supervision is due to class size, it does not contradict the previously reported failure to observe a difference in the frequency of supervision as a teacher's activity, since supervision could be reported as a teacher's activity while the pupils were working on an assignment without being one of the three most frequent teachers' activities.

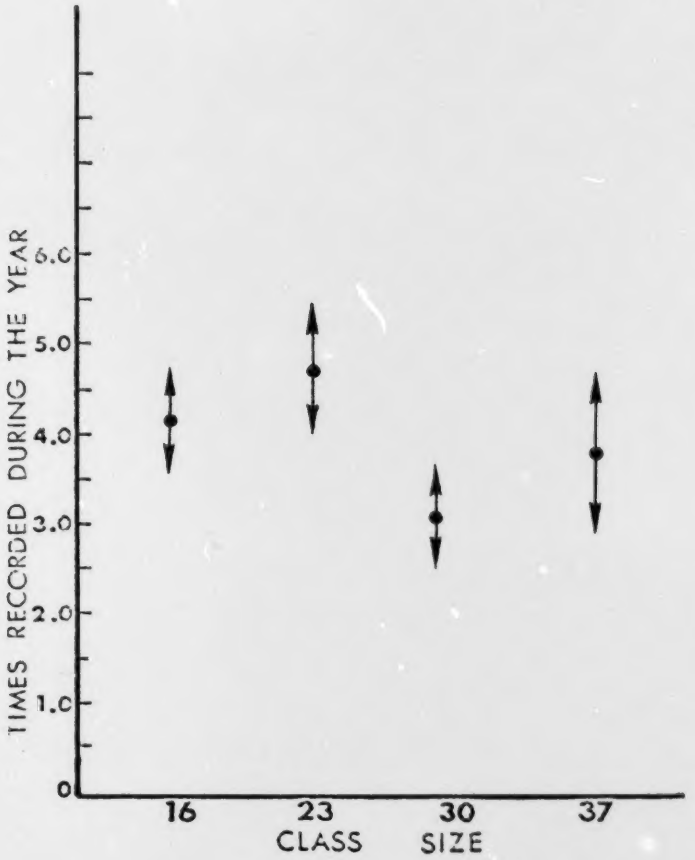


Figure 3. Mean number of observations during which lecturing was one of the three most frequent methods of instruction (95 per cent confidence intervals).

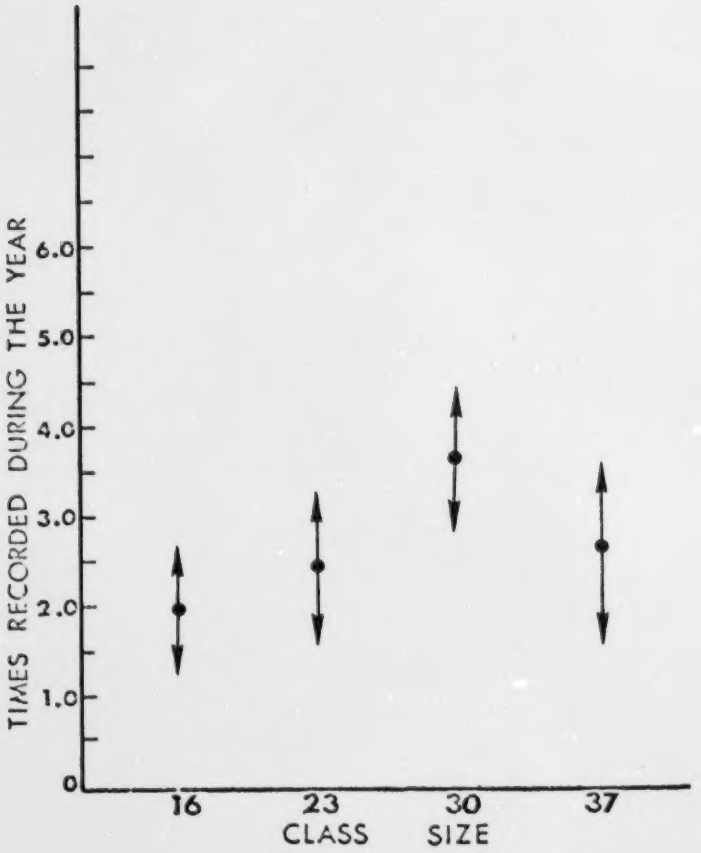


Figure 4. Mean number of observations during which it was reported that the teacher supervised while pupils were working (reliability could not be demonstrated; 95 per cent confidence intervals).

Subject Emphasis. In one section of the TCOS the observer lists the subjects studied during the half-day of observation. Only three subject categories appeared frequently enough to be worth analysing; not surprisingly, they were reading, mathematics and all language arts subjects (including reading). The reading and mathematics variables were assessed by counting the number of observation visits during each year when the subject was reported as being taught in the class, so that the maximum value of each variable was eight. The language arts variable was the mean number of language arts subjects reported during an observation; the maximum value was fifteen, as there are fifteen language arts categories in the coding systems. The mathematics and language arts variables were not affected by class size. Reading was the only reliable variable and was affected by class size. It was reported significantly more frequently in classes of 16 than in classes of 30; no other differences were significant. The linear trend accounted for 66.6 per cent of the variance between cells. See Table 19 and Figure 5 for results. An analysis of variance of pseudo-values calculated by Mosteller and Tukey's jackknife procedure approached significance; the probability of the F value was .054.*

Physical Conditions. The physical characteristics of the classroom investigated by the TCOS are furniture arrangement (that is, desk arrangement) and noise. Furniture arrangement was classified as audience (that is, all desks facing in the same direction), less than

---

* It was originally intended that the amount of time spent on each subject be analyzed as well, However, a satisfactory measure could not be devised. It is difficult, for example, to decide how much time is spent on each subject when the teacher puts on the board a list of work to be done during the rest of the day and the pupils spend the rest of the day working on subjects in which their work is not finished.

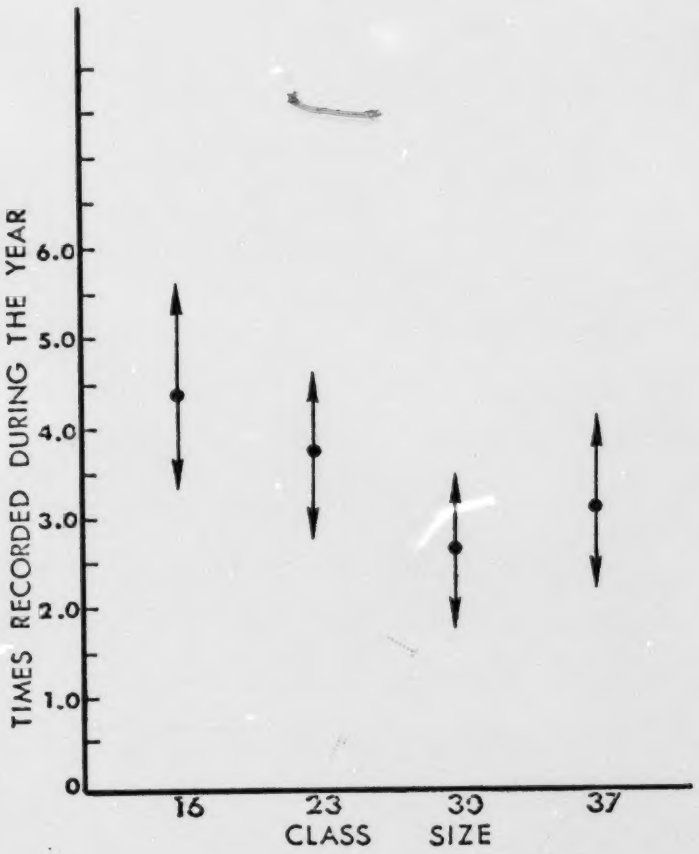


Figure 5. Mean number of observations during which reading was taught (95 per cent confidence intervals).

100 per cent audience but at least 50 per cent, face-to-face groups, or no pattern (this could be expected when furniture is pushed to one side for special reasons). The variables were the numbers of times each arrangement was reported during the year in each class, so that the maximum value of each variable was eight. All the variables were reliable and not affected by class size. The classroom noise variables were similarly constructed. Classroom noise was classified as quiet, a general hum of activity, noisy or widely varying. The counts of quiet and of noisy classes were reliable and not affected by class size. The reliability of the other two noise variables could not be demonstrated and they also were not affected by class size.

Use of Educational Resources. In one section of the TCOS the observer checks a list of educational aids to indicate which were present and which were used. Three categories of educational aids are investigated: audio-visual, written and mechanical-concrete.

The audio-visual aids investigated were: radio, television set, record player, film projector, film strip or slide viewer.* The variables were the total number of aids listed as present during a year and the total number which were used. The first variable was reliable, but the reliability of the second could not be demonstrated, probably because audio-visual aids were rarely used (about once in every five observations). Neither variable was affected by class size. The proportion of aids available that were used was not affected by class size.

Written aids investigated were: chalkboard, textbooks, reference material, globe or map, signs or charts. The variables assessed were similar to the audio-visual variables. Again, the count of aids present was reliable while the count of aids used was not.

---

* Other aids were listed on the TCOS form in the Technical Supplement but only aids which were listed in both years of the study were investigated.

The number of written aids present was not affected by class size, but the number of written aids used was. Written aids were used more frequently in classes of 23 and 30 than in classes of 16; no other differences were significant. The linear trend was not significant. As would be expected, the proportion of available written aids which were used was related to class size in the same way. See Table 19 and Figure 6 for results. As the count of aids used was not demonstrably reliable, the observed difference is not easily attributed to class size.

The mechanical-concrete aids investigated were: games, pictures, animals, plants, typewriter, specimen objects, models, art supplies, piano, other musical instruments and scientific equipment. The variables assessed were similar to the audio-visual variables and only the count of aids present was reliable. Neither of the variables was affected by class size.

In another section of the TCOS, the observer estimates the proportion of the class which used each type of aid. The variables derived from these estimates were the number of times in a year that no pupils used each type of aid, the number of times that some used each type of aid, and the number of times that all used each type of aid. None of these variables was demonstrably reliable, and none was affected by class size.

Classroom Atmosphere. The set of items used to assess classroom atmosphere was scored by finding the number of items on which the observer agreed with a description of "good" atmosphere or disagreed with a description of "bad" atmosphere. The set of items was reliable. A factor analysis of the items was performed after the first year of the study and extracted eleven factors which explained 61.5 per cent of the variance. The three most important factors accounted for 36.0

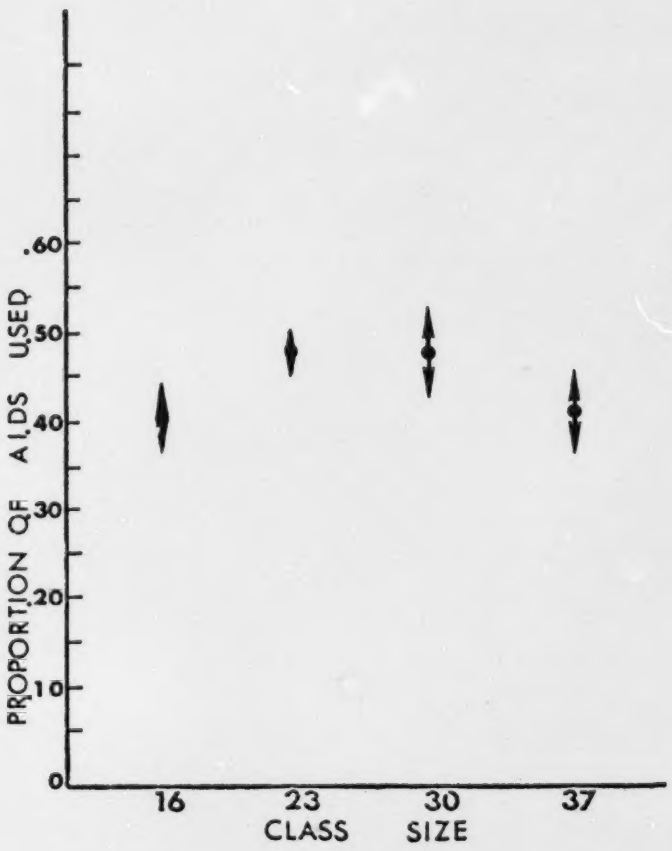


Figure 6. Proportion of available written aids used. (Reliability could not be demonstrated; means and 95 per cent confidence intervals.)

per cent of the variance. Items which loaded highly on the three largest factors appeared to refer to the friendliness of teacher-pupil interaction, the teacher's control of the classroom, and the use of emotional, authoritarian methods by the teacher. The Pearson correlation coefficient for the relationship between teachers' scores in the two years was .86.

Classroom atmosphere was not affected by class size.

Indicators of Quality

Three Indicators of Quality scores were investigated -- the number of positive signs recorded (the positive score), the number of negative signs recorded (the negative score) and the result of subtracting the negative score from the positive score (the difference score). The sum of the scores from all five observations of a teacher during a year was reliable. The split-half reliability coefficients for positive scores were .76 in the first year and .80 in the second year; for negative scores the coefficients were .84 in the first year and .79 in the second (there are no significant differences between any of these coefficients). Correlations between years were also significant.

Since the negative scores were significantly lower in the first year of the study than in the second, the first year's scores were adjusted and the adjusted negative scores were used to calculate the difference scores.

Indicators of Quality scores were not affected by class size.

An Alternative Analysis of the Observational Data

One problem with the preceding analysis of the observational data is that the power of the statistical test is rather low.*

* "The power of a statistical test is the probability that it will yield statistically significant results" (Cohen, 1969).

For example, if the variance of class size means is moderate and the significance criterion is a chance probability less than .05, the power of the one-way analysis of variance to detect a difference of one standard deviation between the highest and lowest class size means is about 70 per cent (all power estimates in this section are taken from Cohen, 1969).

This degree of power is lower than Cohen's (1969) suggested scientific standard of 80 per cent, even though a difference of one standard deviation is large. The power of the one-way analysis of variance could have been increased, but only by raising the significance criterion. However, it was possible to devise a more powerful analysis of observational data.

Twenty-eight teachers had both a small class (that is, a class of 16 or 23) and a large class (a class of 30 or 37) during the study. By comparing their scores in their larger and smaller classes with a t-test, much smaller differences could be detected. For example, if there is no correlation between scores in the two years, this analysis has power of 80 per cent to detect a difference of .8 standard deviations at the .05 level. If there is a correlation of .6 (a frequently observed degree of correlation in the present study) the test has power of 80 per cent to detect a difference of half a standard deviation. This increase in power is due to the increase in size of the groups compared, the reduction in the number of groups compared, and the finer control of effects due to teacher. This analysis may be less interesting to interpret because information is not provided for all the class sizes studied; this means, among other things, that certain non-monotonic effects will not be detected. As well, small significant differences may not be very important.

All tests were two-tailed. Sixty-three variables were analyzed.

Toronto Classroom Observation Schedule. This analysis detected five additional differences in the TCOS variables (see Table 20). The frequencies of neutral verbalizations about pupils' behaviour and of all verbalizations about pupils' behaviour were higher in the smaller classes, as were scores for verbal participation. Smaller classes were more likely to be described as quiet and also were less likely to be described as noisy. These variables were all reliable.

Indicators of Quality. Two additional differences were detected. Difference scores were higher and negative scores were lower in the smaller classes. As well, the value of t for positive scores was marginal ($t = 1.92$, $p = .07$). As can be seen in Table 20, both significant differences were quite small relative to the standard deviations of the comparison groups. As well, it should be remembered that the Indicators of Quality scores are the sums of the scores from five observations, so that the difference in negative scores is equivalent to a difference of .73 at each observation, and the difference in the difference scores is equivalent to a difference of 1.44 at each observation.

Student Opinions and Attitudes

Attitudes toward School Scale

Total Scale Analysis. This "Attitudes toward School" scale contained 30 items (18 positively worded, 12 negatively worded) with which students could either "agree," "disagree" or be "undecided."

TABLE 20

ADDITIONAL SIGNIFICANT DIFFERENCES IN OBSERVATIONAL VARIABLES DETECTED BY THE t-TEST

Area of Effect	Variable	Small Class Mean (and Standard Deviation)	Large Class Mean (and Standard Deviation)	Difference* between Means	Standard Error	t**
Teacher-pupil interaction	TVB - all verbalizations about pupils' behaviour maximum = 32	1.11 (0.77)	1.60 (1.32)	-0.49	0.19	-2.63
Teacher-pupil interaction	TVB - all neutral verbalizations about pupils' behaviour maximum = 32	0.78 (0.52)	1.15 (0.78)	-0.38	0.13	-2.90
Pupil participation	IPP - all verbal participation maximum = 32	1.14 (0.39)	0.97 (0.39)	0.18	0.08	2.31
Physical conditions	Classroom noise -- quiet -- maximum = 8	3.14 (2.26)	2.32 (2.04)	0.82	0.31	2.66
Physical conditions	Classroom noise -- noisy -- maximum = 8	0.71 (1.24)	1.57 (1.99)	-0.86	0.285	-3.01
Indicators of Quality	Difference score	9.30 (19.80)	2.09 (21.15)	7.21	2.87	2.51
Indicators of Quality	Negative score	24.34 (11.49)	27.98 (12.16)	-3.64	1.723	-2.11

* Obtained by subtracting the mean of the large class from the mean of the small class.

** At the .05 level, the critical value of t is ± 2.052.

In obtaining a total scale score, all items were scored in the positive direction (i.e. "agree" with a positively worded item, or "disagree" with a negatively worded item with the "undecided" responses omitted from the analysis). The following procedures were followed in analyzing the data: first, a one-way analysis of variance by class size was conducted; second, the variability due to two variables, year of the study (i.e. Grade 4 and Grade 5) and teacher, was removed using a multiple linear regression technique; third, the analysis of variance was repeated on the "residual" data. The class was the unit of analysis in all cases.

The mean total scale scores (unadjusted and residuals) are presented by class size (Table 21) and they are plotted with the 95 per cent confidence intervals (Figure 7).

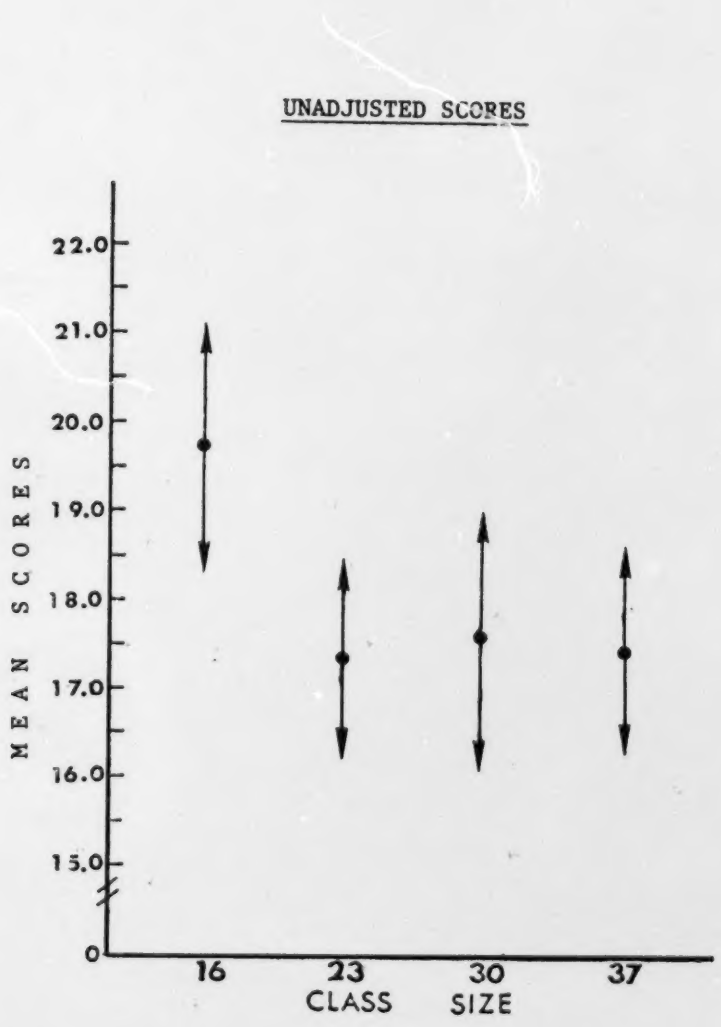
TABLE 21

RESULTS OF PUPILS' ATTITUDES TOWARD SCHOOL: TOTAL SCALE SCORES

	Class Size				Anova	
	16	23	30	37	df	F
Unadjusted Mean	19.71	17.35	17.57	17.48	3,58	3.65*
(Standard Deviation)	(2.50)	(2.08)	(2.62)	(2.14)		
Residual Mean	0.29	0.11	0.11	-0.08	3,58	1.08
(Standard Deviation)	(0.29)	(0.66)	(0.81)	(0.39)		

* $p < .05$

First, the analysis of variance on the unadjusted means revealed a significant effect apparently due to class size, $F(3,58) = 3.65$, $p < .05$. Pair-wise contrasts revealed that attitudes toward school were significantly more positive for students in size 16 than for their peers in size 23, in size 30 and in size 37 respectively; $t(58) = 2.5$, $p > .001$ in each case. No other pair-wise contrasts were significant.



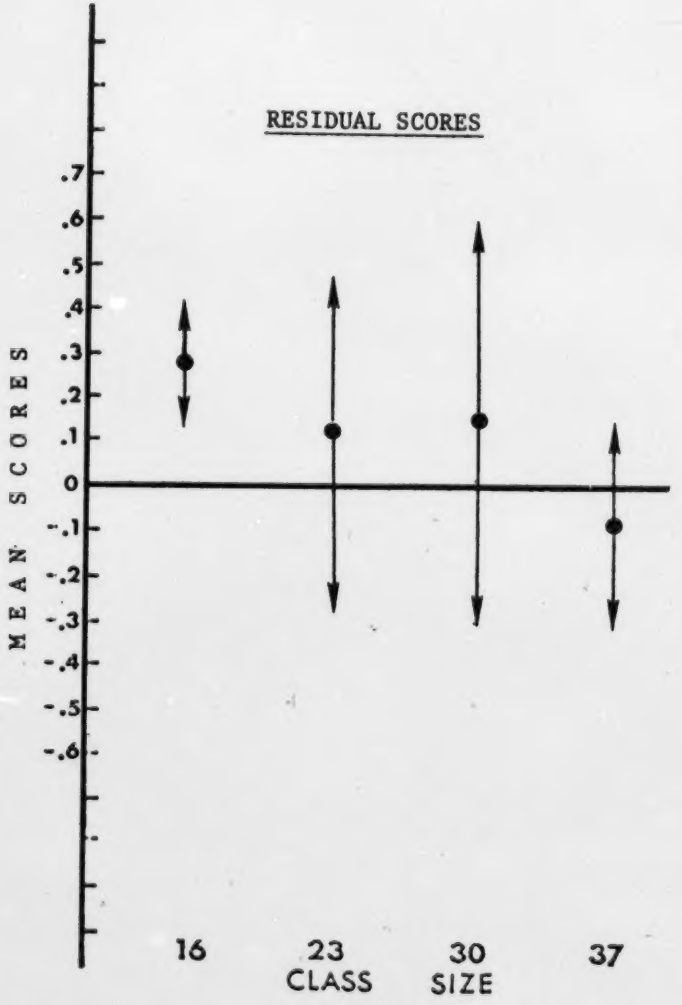


Figure 7. Pupils' Attitudes toward School -- total scale: means and 95 per cent confidence intervals for both unadjusted scores and residuals by class size, based on class as unit of analysis.

However, when this analysis was repeated on the residual data after removing the variability due to year and teacher, there no longer was a significant effect due to class size, $F\ (3,58) = 1.08$, $p > .05$. Thus, the apparent class size effect noted initially was an artifact.

Discriminant Scale Analysis. The 30 items of the Attitudes toward School scale were subjected to a discriminant analysis. This method first selects the item which is most useful in discriminating between students in the four class size groups. In a step-wise procedure a second discriminating item is selected as best able in combination with the first item to improve the discrimination between groups. This procedure is continued until all the items have been entered or until selecting a new item does not significantly increase the ability to discriminate among the groups.

This analysis resulted in the selection of the thirteen items (listed in Table 22) which were used to form a discriminant scale score. A one-way analysis of variance was performed by class size (on both unadjusted and residual data) following the same procedures as used for the total scale.

The results (in Table 23 and Figure 8) revealed a significant effect apparently due to class size, $F\ (3,58) = 5.80$, $p < .01$. On repeating the analysis of variance with the residuals (i.e. after removing the variability due to teacher and year) the difference does not remain $F\ (3,58) = .099$. As above, the apparent class size effect on student attitudes was misleading.

TABLE 22

ITEMS ON PUPILS' ATTITUDES TOWARD SCHOOL
SCALE SELECTED BY THE DISCRIMINANT ANALYSIS

Step Number	Item on Attitude Scale	Wilks' * Lambda
1	My classroom is crowded.	.818
2	I am happy in my classroom.	.801
3	Everyday we do the same old thing.	.791
4	The books we read are boring	.783
5	There are too many rules in my class.	.778
6	Pupils in my class are so noisy I can't do my work.	.775
7	In our classroom we don't have enough interesting things like plants and animals.	.771
8	My teacher has time to talk to me about my work.	.767
9	In my class we have time to discuss new ideas.	.764
10	My teacher knows me very well.	.761
11	I wish we could do more new things in our class.	.759
12	I can see the work on the blackboard from my own seat.	.757
13	The teacher never seems to ask me to do anything special.	.754

* $p < .001$ for all coefficients.

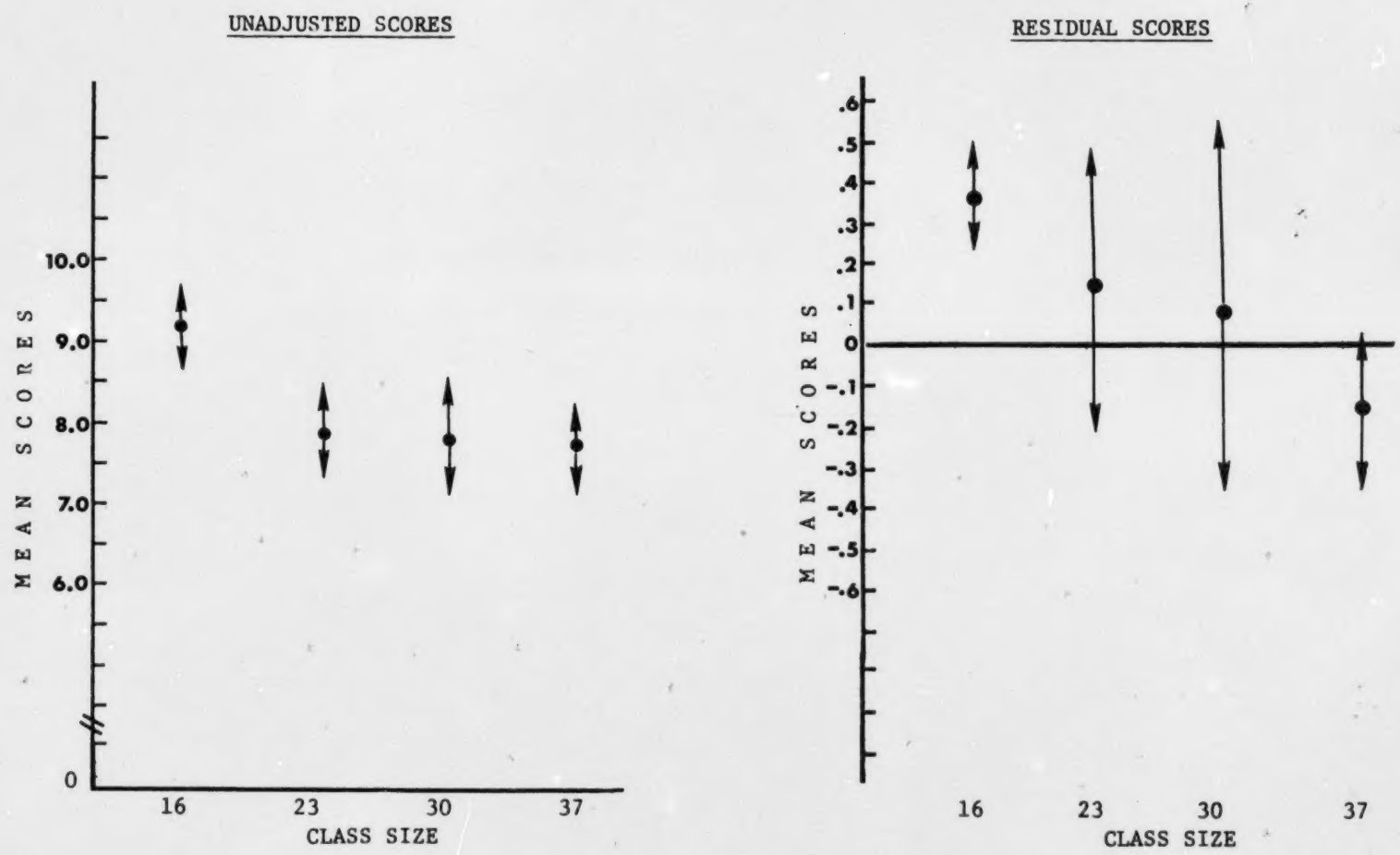


Figure 8. Pupils' Attitudes toward School -- Discriminant Scale: Means and 95 per cent confidence intervals for both unadjusted scores and residuals by class size.

TABLE 23

RESULTS OF PUPIL ATTITUDE TOWARD SCHOOL:
DISCRIMINANT SCALE SCORES

	Class Size				Anova	
	16	23	30	37	df.	F
Unadjusted Mean	9.18	7.94	7.87	7.76	3,58	5.80*
(Standard Deviation)	1.07	1.08	1.27	.96		
Residuals Mean	.37	.14	.10	-.15	3,58	2.18
(Standard Deviation)	.26	.66	.84	.34		

* $p < .001$.

Students' Opinions of Subject Emphasis

Students were asked whether, in their opinion, the amount of class time spent on three core subjects (Reading, Writing and Mathematics) and on three non-core subjects (Music, Art, Gym) was "too much," "okay" or "too little." The responses to each item are presented graphically by class size (see Figure 9).

For each subject, the largest proportion of students felt that the amount of time spent was "okay." However, it can be observed that there was a tendency for a larger percentage of students to respond that "too much" time was spent on mathematics. In addition, in comparison to the core subjects, there was a tendency for a larger percentage of students to indicate that "too little" time was spent on the non-core subjects.

For each of the six subjects, a mean score was calculated by class size group (coding 1 = "too much;" 2 = "okay;" 3 = "too little") and one-way analyses of variance were performed.

It was only for Art that a significant effect due to class size was found, $F(3, 1467) = 2.88$, $p < .05$. This finding signified

Time Spent On	Class Size	Per Cent of Students Responding		
		Too Much	Okay	Too Little
Reading	16	12.0	75.5	12.4
	23	12.2	73.3	14.5
	30	13.2	73.1	13.7
	37	10.1	76.5	13.4
Writing	16	11.7	72.1	16.2
	23	13.6	72.6	13.9
	30	14.1	64.1	21.8
	37	12.0	69.6	18.4
Mathematics	16	21.9	65.3	12.8
	23	30.6	59.4	10.0
	30	26.9	57.2	15.9
	37	24.7	64.7	10.6
Music	16	11.6	50.8	37.6
	23	6.8	54.3	38.9
	30	11.2	40.1	48.7
	37	8.2	54.9	36.8
Art	16	12.4	60.7	26.9
	23	9.4	60.3	30.3
	30	10.6	53.6	35.8
	37	6.7	59.0	34.3
Gym	16	7.5	51.9	40.7
	23	4.1	54.7	41.2
	30	4.0	47.6	48.4
	37	5.4	53.1	41.5

Figure 9. Students' opinions about the amount of time spent on various subjects.

that in comparison to students in class size 16, students were significantly more likely to feel that too little time is spent on Art if they were in class size 30 ($t(1464) = -2.19$, $p < .05$) and if they were in class size 37 ($t(1464) = -2.77$, $p < .005$).

Student Response to Semantic Differential

Students' attitudes toward the concept "My Classroom" were assessed through a semantic differential technique. A score was created from the information provided by the students' responses to the fifteen adjective pairs. The method of obtaining this score was as follows:

(a) The five point scale representing a continuum between the polar adjectives was reduced to a dichotomous variable having a value of 1, if the student expressed a positive attitude on the scale, and 0 otherwise;

(b) A score for each student was calculated by counting the number of positive (1) values;

(c) A mean score for each class was calculated by summing each students' score and dividing by the number of students in the class;

(d) Finally, a mean score for each class size was computed. The means and standard deviations are presented in Table 24 below. The possible range of values is from zero to fifteen where the higher score value indicates a more positive attitude toward "My Classroom."

An analysis of variance revealed no significant difference between class sizes, $F(3,58) = 0.46$. To parallel the procedure used in the analysis of the data from the other attitude instrument, regression analysis was used to remove teacher and year effects and the analysis of variance was repeated using the residuals. Again, no difference was found between class sizes, $F(3,58) = 0.72$.

TABLE 24

MEANS AND STANDARD DEVIATIONS FOR SCORES DERIVED FROM PUPILS' SEMANTIC DIFFERENTIAL "MY CLASSROOM"

Scores	Class Size				Anova	
	16	23	30	37	df	F
Total score mean	8.63	8.93	8.78	9.2	3,58	0.46
(Standard deviation)	1.20	1.41	1.96	1.39		
Residuals total score	0.03	0.18	0.26	0.06	3,58	0.73
(Standard deviation)	0.39	0.71	0.75	0.28		
Discriminant score means	6.16	6.44	6.36	6.77	3,58	0.51
(Standard deviation)	0.93	0.98	1.48	1.02		
Residuals discriminant score	0.06	0.14	0.28	0.10	3,58	0.92
(Standard deviation)	0.41	0.72	0.74	0.28		

The fifteen dichotomized variables (created as described above) were entered in a discriminant analysis. Eleven of the items were selected as most able, when taken in linear combination, to contribute to discrimination between responses of students in the four class sizes. A discriminant scale score was formed by these eleven items (means and standard deviations for this scale are also found in Table 24 above), and the analysis of variance was repeated. No difference between class sizes was found, $F(3,58) = 0.51$. Neither was any significant difference found in an anlysis of variance performed with the residuals for the discriminant scale, $F(3,58) = 0.92$.

Student Outcome Measures

The effects of class size on student outcome measures were assessed using six student performance measures: art, composition, reading, vocabulary, mathematics-concepts and mathematics-problem solving, and also one affective measure: self-concept. For each measure the following procedures were followed:

(a) What Students Were Included in the Analyses? The data analyses included only those students who were present at the beginning of the school year (i.e. end of September) and who were still present in the class at the time the particular measure was collected (see Timetable of Data Collection Activities, Table 11, p. 50). Thus, students recently transferring into a particular class are not included and those who transferred out of the classes involved in the study were not followed up for any measures.

(b) How Were the Data Analyzed? The variability due to two variables that had to be controlled, year of the study (i.e. Grade 4 vs Grade 5) and teacher, were first removed using a multiple linear regression technique. A one-way analysis of variance by class size was then conducted on the "residual" data. The class was the unit of analysis in all cases. In an initial examination of the data there were great differences, both among schools and among teachers. The procedure of taking out the effects by year and by teacher, as described above, is a technique for dealing with the complexity of the data.

The mean scores (unadjusted values and residuals) are presented by class size in Table 25 for each measure. The residual means and 95 per cent confidence intervals* are plotted for each class size (see Figures 10 to 13). The results revealed:

(a) no significant differences attributable to class size for art ($F_{3,58} = 1.29$), composition ($F_{3,57} = 0.88$), vocabulary ($F_{3,58} = 0.71$), reading ($F_{3,58} = 0.16$), mathematics-problem solving ($F_{3,58} = 2.58$) and self-concept ($F_{3,58} = 1.38$); $p < .05$ in each case.

(b) for mathematics-concepts, there was an overall significant effect due to class size (see Figure 13) $F_{3,58} = 4.11$, $p < .01$.

In order to pinpoint this effect, pair-wise contrasts were conducted: students in size 16 performed significantly better in mathematics-concepts than their peers in size 30 ($t_{58} = 3.16$, $p < .005$) and than their peers in size 37 ($t_{58} = 2.87$, $p < .01$). None of the other pair-wise contrasts were statistically significant.

---

* See footnote on page 72 for an explanation of confidence intervals.

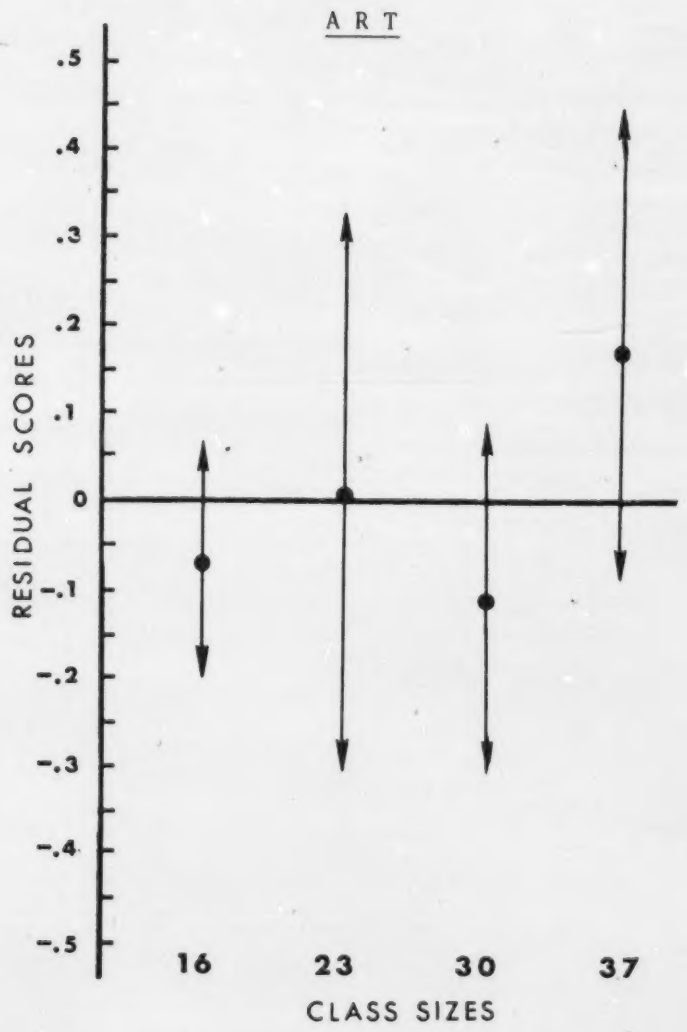
TABLE 25

RESULTS OF STUDENT OUTCOME MEASURES BY CLASS SIZE

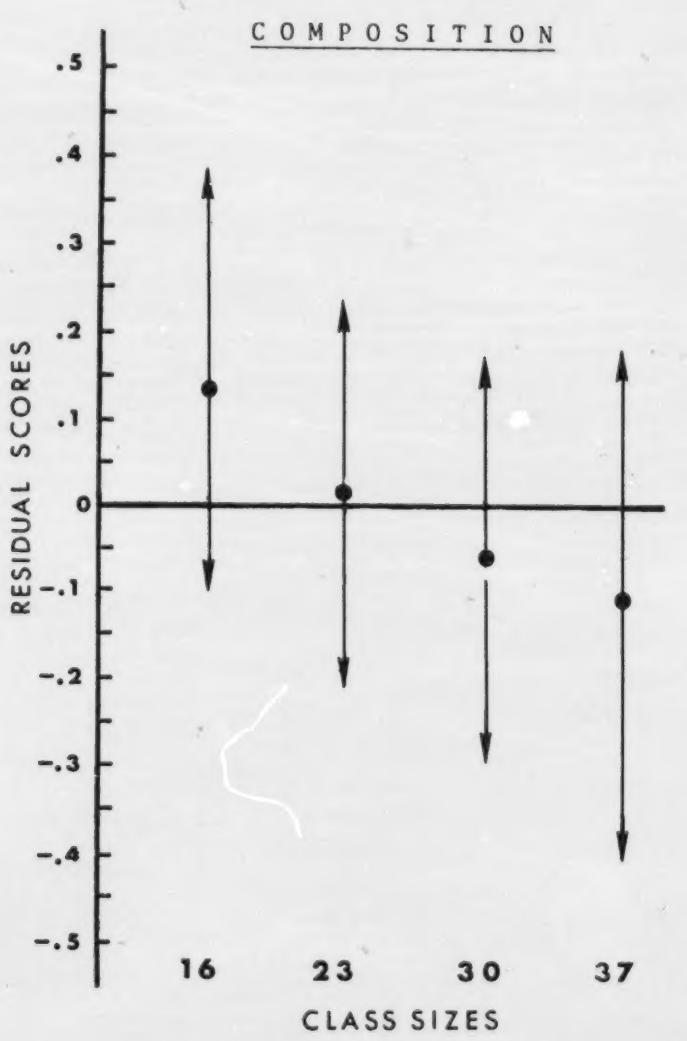
Measure	Class Size 16	23	30	37	Anova * df	F
Art						
unadjusted mean	6.47	6.30	6.49	6.58		
(standard deviation)	(0.39)	(0.91)	(0.58)	(0.50)		
residual mean	-0.07	0.01	-0.11	0.18	3,58	1.29
(standard deviation)	(0.24)	(0.59)	(0.35)	(0.49)		
Composition						
unadjusted mean	4.56	4.40	4.51	4.46		
(standard deviation)	(0.77)	(0.85)	(0.66)	(1.01)		
residual mean	0.14	0.02	-0.06	-0.11	3,57	0.88
(standard deviation)	(0.46)	(0.43)	(.043)	(0.51)		
CTBS Vocabulary						
unadjusted mean	43.16	43.58	42.42	42.84		
(standard deviation)	(9.55)	(9.97)	(6.51)	(8.06)		
residual mean	-0.00	0.04	0.03	-0.07	3,58	0.71
(standard deviation)	(0.22)	(.024)	(0.24)	(0.14)		
CTBS Reading						
unadjusted mean	44.08	44.47	44.05	44.28		
(standard deviation)	(7.99)	(8.62)	(6.00)	(7.22)		
residual mean	-0.00	-0.04	0.02	0.02	3,58	0.16
(standard deviation)	(0.30)	(0.30)	(0.27)	(0.25)		
CTBS Math - Concepts						
unadjusted mean	47.41	45.60	44.03	44.69		
(standard deviation)	(9.24)	(8.52)	(6.27)	(7.36)		
residual mean	0.15	0.00	-0.09	-0.07	3,58	4.11**
(standard deviation)	(0.22)	(0.20)	(0.18)	(0.23)		
CTBS Math - Problem Solving						
unadjusted mean	51.17	49.33	48.20	48.82		
(standard deviation)	(8.83)	(8.21)	(6.36)	(6.84)		
residual mean	0.14	0.04	-0.09	-0.10	3,58	2.57
(standard deviation)	(0.27)	(0.28)	(0.27)	(0.30)		
Self-Concept						
unadjusted mean	20.16	20.34	20.21	20.25		
(standard deviation)	(2.42)	(1.91)	(1.74)	(2.15)		
residual mean	0.09	-0.08	0.13	-0.14	3,58	1.38
(standard deviation)	(0.50)	(0.35)	(0.35)	(0.51)		

* Only data for the analyses of variance on the residual means are presented here. Similar analyses were concluded on the unadjusted means but none of these were significant at the .05 level.

** $p < .01$

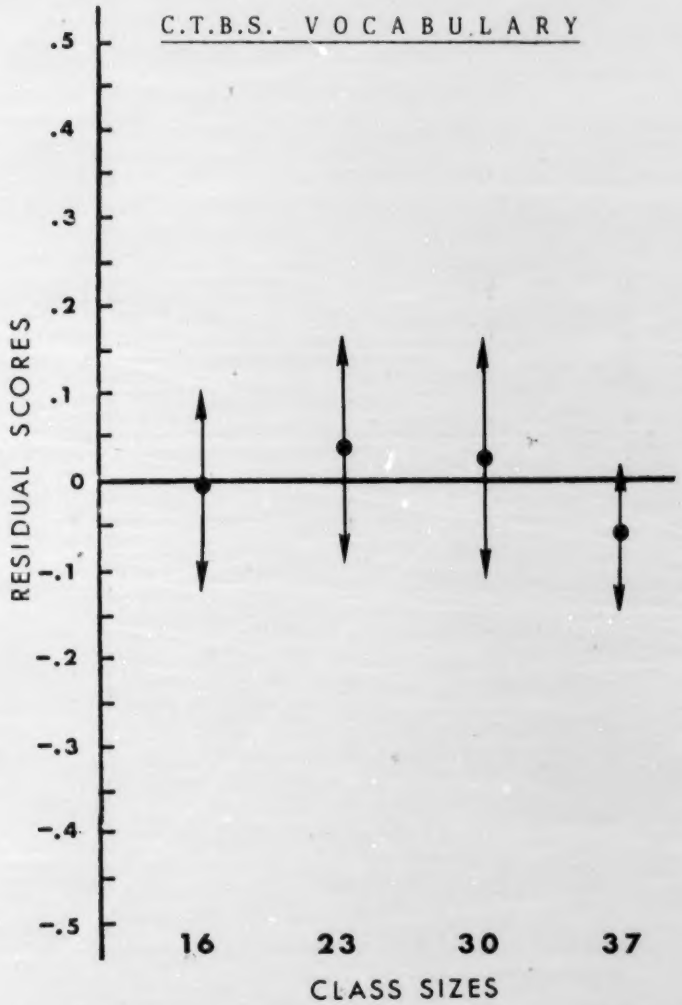


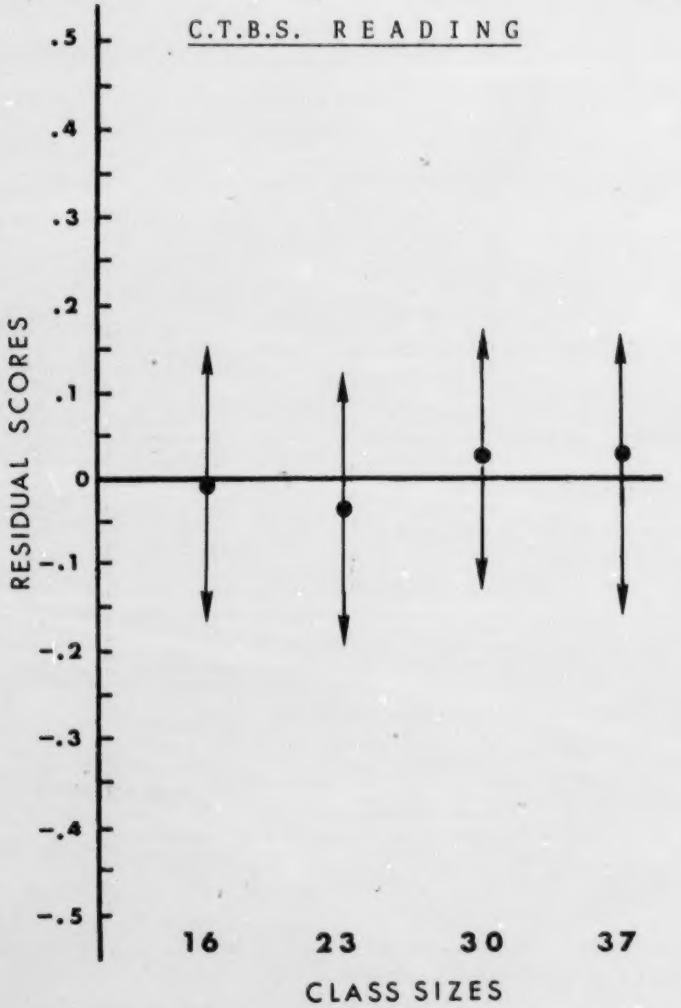
123



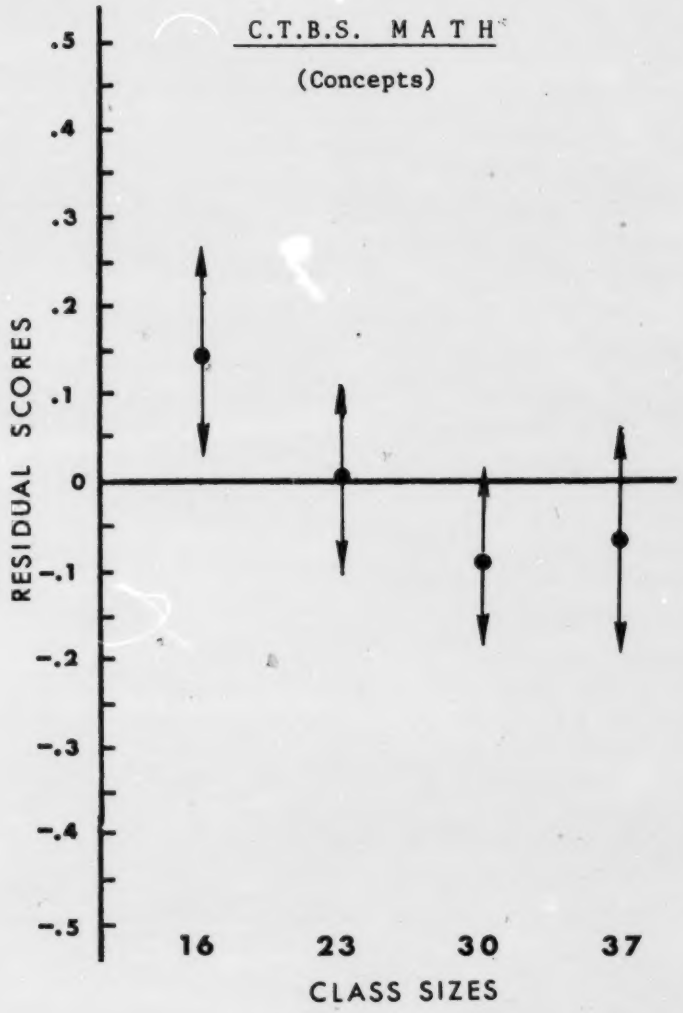
124

Figure 10. Student Performance Measures: Residual Means and Confidence Intervals (95%) by Class Size.





125 Figure 11. Student Performance Measures: Residual Means and Confidence Intervals (95%) by Class Size. 126



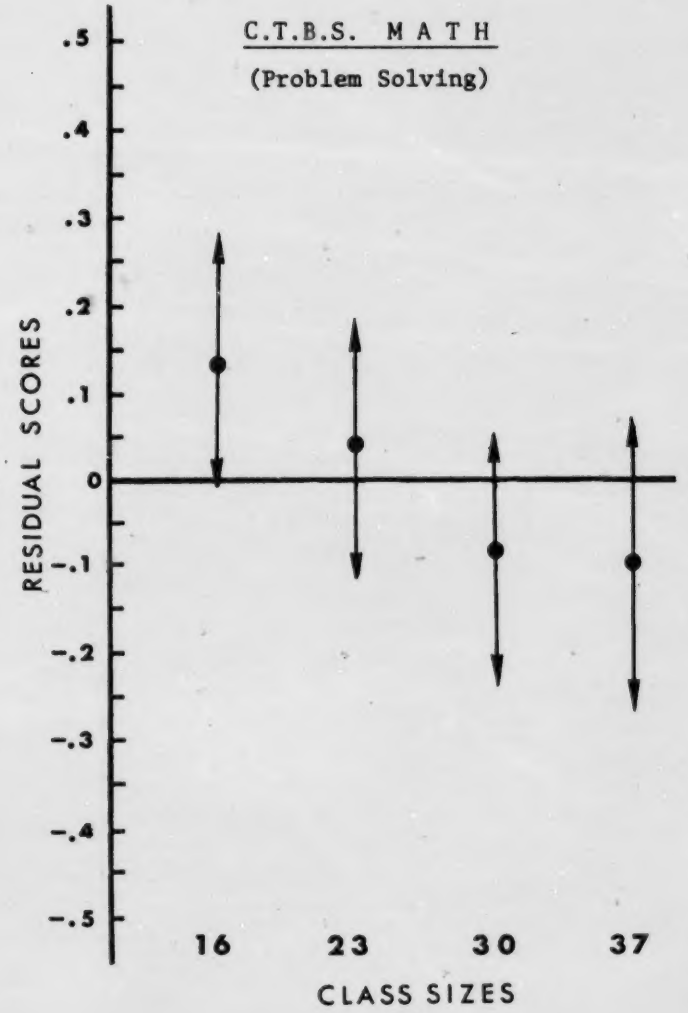


Figure 12. Student Performance Measures: Residual Means and Confidence Intervals (95%) by Class Size.

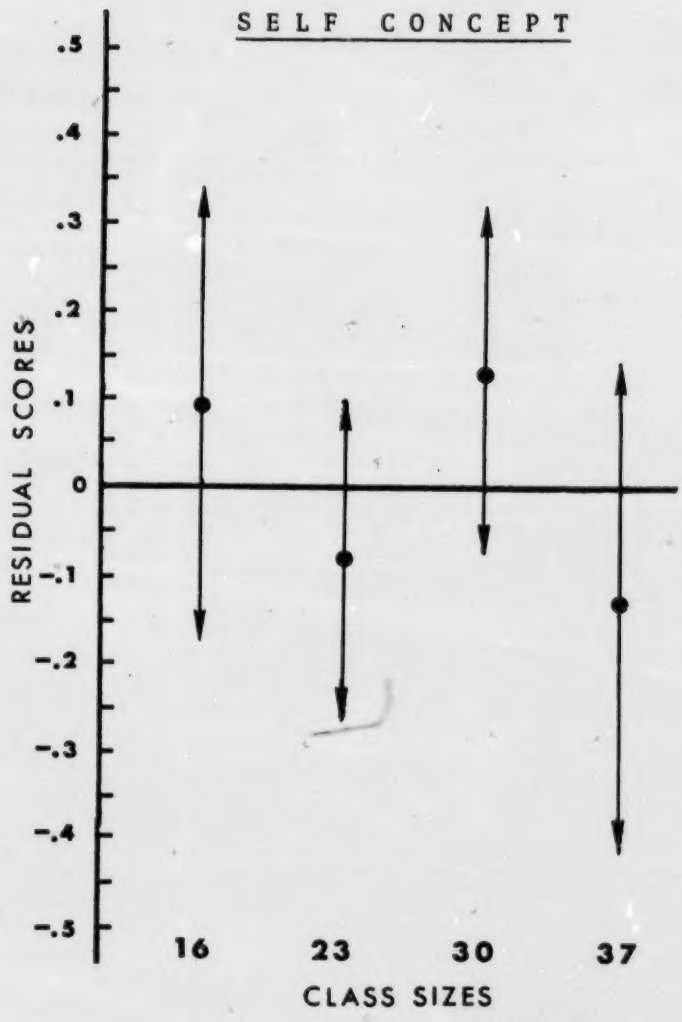


Figure 13. Students' Self-Concept: Residual Means and Confidence Intervals (95) by Class Size.

For mathematics-problem solving, the results, while not significant, followed a pattern that was similar to mathematics-computation.

Parental Opinions

During the Spring of the second year a questionnaire was sent out to parents of students who were enrolled in the study. Returns were received from 500 parents, a response rate of 62 per cent. The response rate by class size is presented in Table 26.

TABLE 26

RESPONSE RATE TO THE PARENT QUESTIONNAIRE

Class Size	Number Distributed	Number Returned	Response Rate
16	152	116	76.3%
23	209	114	54.5%
30	204	112	54.9%
37	245	158	64.5%
TOTAL	810	500	62 %

Parental Awareness of Child's Class Size

Parents were asked to indicate the number of students in their child's class, for the current school year (i.e. the second year of the study) as well as for the previous year (i.e. the first year of the study). A correct estimate was defined as one within the interval bounded by the child's assigned class size ±3 students.

The majority of parents (71 per cent) correctly estimated the size of their child's current class. However, when parents

attempted to estimate the previous year's size, accuracy decreased to 55.6 per cent.

Parental Opinion about the Size of Child's Class

Parents were asked to indicate whether the number of students in their child's class for the current year was "too large," "too small" or "about right." The pattern of responses varied significantly according to the child's assigned class size, (χ^2 (9) = 124.41, $p < .001$); see Table 32 at the end of this section for a summary of those frequency distributions that are significant.

There was a tendency for parents whose children were in the larger sizes (i.e. 30, 37) to rate them "too large" (57 per cent of these parents' responses) and for parents of children in the smaller class sizes (i.e. 16, 23) to rate them "about right" (75 per cent of parents). For the group of parents (i.e. 71 per cent of the sample) who had accurately estimated the size of their child's current class:

-- 68 per cent having children in size 37 rated it as "too large,"

-- 84.1 per cent having children in size 16 felt that it was "about right."

Parental Priorities for Improving Education

Parents were asked to rate seven possible changes in education as to whether they would give them a high or low priority in terms of improving their children's education. The responses are presented and ranked in Table 27.

These findings revealed:

-- at least 50 per cent of the respondents indicated that each item was of "high priority" in improving education;

-- parents were most likely to assign high priority to "reducing the size of classes," followed by "improving the quality and quantity of teaching materials";

-- parents were least likely to assign high priority to "improving existing school facilities" and "offering a wide variety of classes in courses such as art, music, drama."

TABLE 27

PARENTAL PRIORITIES FOR IMPROVING EDUCATION

Rank	Item	% High Priority Responses
1	Reducing the size of classes	72.7
2	Improving the quality and quantity of teaching materials	70.9
3	Improving the qualifications of teachers	69.4
4	Improving the communication between the school and the home	68.7
5	Increasing the number of teachers	68.3
6	Offering a wide variety of classes in courses such as art, music, drama	53.7
7	Improving existing school facilities	53.2

Parental Opinions about What Class Size Affects

Parents were asked to respond ("yes," "no") to the ten statements of "I think class size affects ___" (see Table 28).

A large proportion of parents held the opinion that class size affects "student-teacher relationships," "discipline," "pupils' interest," "quality of teaching" and "teaching methods." A smaller proportion of parents felt that class size affects "parental participation" and "communication between the home and school." It was also observed that "student-teacher relationships" was the item most frequently cited by parents of children in both the two smaller sizes,

(16 and 23) while "discipline" was most frequently mentioned by parents of children in both the larger sizes (30 and 37).

TABLE 28

PARENTAL OPINIONS AS TO WHETHER SELECTED VARIABLES ARE AFFECTED BY CLASS SIZE

Rank	Item	% "Yes" Responses
1	student-teacher relationships	87.5
2	discipline	86.8
3	pupils' interest	84.7
4	the quality of teaching	84.6
5	teaching methods	82.7
6	the cost of education	70.9
7	curriculum	63.6
8	extracurricular activities	60.7
9	communication between home and school	50.9
10	parental participation	44.7

It is most important to note that 29 per cent of the parents did not think that class size affects the cost of education.

Parents' response patterns varied significantly according to their child's class size for quality of teaching (χ^2 (3) = 8.23, $p < .05$) and extracurricular activities (χ^2 (3) = 8.43, $p < .05$). The percentage of "yes" responses to these two items tended to be higher for parents of children in sizes 16 and 30 than for parents of children in sizes 23 and 37.*

* Refer to the summary of these distributions at the end of this section.

Parental Perception of Child's Improvement

Parents were asked to respond ("yes" or "no") to the statement "during the current school year, my child has improved in" for a twelve item list (see Table 29).

The percentage of parents answering "yes" to each item tended to be high (ranging from 56.7 per cent to 90.8 per cent).

The five skills in which parents were most likely to feel that their child had improved included basic skills (reading, speaking, and mathematics), along with interpersonal and self-skills (getting along with others and self-confidence). Parents were least likely to feel that their child had improved in musical ability.

TABLE 29

"DURING THE CURRENT SCHOOL YEAR, MY CHILD HAS IMPROVED IN"

Rank	Item	% Yes Responses
1	Getting along with others	90.8
2	Reading	90.6
3	Speaking	89.8
4	Mathematics	88.4
5	Self-confidence	87.0
6	Physical skills	84.8
7	Work habits	80.9
8	Manners	80.1
9	Creativity	79.8
10	Handwriting	77.5
11	Art work	77.3
12	Musical ability	56.7

Parental responses to these skills were examined as a function of their child's class size, the response pattern tended not to vary with class size (χ^2 analyses, $p > .05$). For mathematics there seemed to be a tendency for more parents to note an improvement in mathematics the smaller their child's class size, however this is not significant, χ^2 (3) = 7.76, $p = .052$ (see Table 31).

Parents' Open-ended Comments of Effects of Class Size

Parents also were asked to comment on how they thought the class size study had affected their child. The summary of these comments is presented in Table 30 categorized according to tone (i.e. positive or negative) and according to content (i.e. dealing with effects of class size on student outcomes, on process/environment variables, on teacher or on program). The distribution of positive and negative comments varied significantly by class size (χ^2 (3) = 121.23, $p < .001$). There were significantly more comments made by parents of children in the smaller class sizes (i.e. 16 and 23) than by parents of children in the large class sizes (i.e. 30 and 37); furthermore, the former made primarily positive comments, the latter negative comments.

It can also be seen that most of the positive comments (of parents with children in sizes 16 and 23) related to class size effects on student outcomes and process/classroom environment variables. The negative comments (in sizes 30 and 37) also were concentrated in these two content areas. Only a small proportion of parents' comments dealt with the effects of class size on the teacher and on the program.

Further Findings

It is interesting to note that the amount of contact parents reported having with their child's teacher did not vary significantly

as a function of class size, (χ^2 (9) = 9.64, $p > .05$). However, there was a significant relationship between class size and parent's perception of whether the class took enough field trips, χ^2 (3) = 9.99, $p < .05$. The percentage of parents responding that their child's class took enough field trips was lowest for class size 30 and highest for sizes 23 and 37 (see frequency distribution in Table 31).

TABLE 30

NUMBER OF PARENTS' COMMENTS BY CLASS SIZE

Content of Comments	Class Size 16	23	30	37
POSITIVE COMMENTS				
(a) Student Outcomes (i.e. academic improvement, more involvement in school, developed self-confidence)	42	30	7	7
(b) Process/Environment (i.e. closer relationships with teacher and classmates, more individual attention, positive atmosphere)	18	30	2	1
(c) Effect on Teacher (i.e. teacher's job more positive	0	5	0	0
(d) Effect on Program (i.e. more extracurricular time, choice of activities)	2	3	0	0
TOTAL NUMBER OF POSITIVE COMMENTS*	62	68	9	8
NEGATIVE COMMENTS				
(a) Student Outcomes (i.e. slower academically, less enjoyment, was bored)	0	0	6	6
(b) Process/Environment (i.e. more distant relationship with teacher and classmates, little individual attention, unpleasant atmosphere)	3	2	14	29
(c) Effect on Teacher (i.e. harder for teachers)	1	1	1	7
(d) Effect on Program (i.e. fewer extracurricular activities)	0	0	2	1
TOTAL NUMBER OF NEGATIVE COMMENTS*	4	3	23	43

* Since a comment could cover more than one content area, the total numbers are greater than the number of respondents making comments.

TABLE 31

FREQUENCY DISTRIBUTIONS FOR SELECTED PARENTAL OPINIONS

Question and Response Category	Class Size 16	23	30	37	df	χ^2
Parent opinion of size (1975-76):					9	124.41**
-- too large	11 (11%)	13	48 (57%)	92		
-- too small	6 (3%)	1	1 (1%)	1		
-- about right	79 (75%)	84	46 (34%)	38		
-- no opinion	13 (10.5%)	10	8 (8.5%)	13		
TOTAL	109	108	103	144		
Class size affects the quality of teaching:					3	8.23*
-- yes	92 (92%)	84 (81.5%)	83 (87.4%)	104 (79.4%)		
-- no	8 (8%)	19 (18.5%)	12 (12.6%)	27 (20.6%)		
TOTAL	100	103	95	131		
Class size affects extracurricular activities:					3	8.43*
-- yes	55 (66.3%)	45 (49.5%)	58 (69.1%)	72 (59.5%)		
-- no	28 (33.7%)	46 (50.5%)	26 (30.9%)	49 (40.5%)		
TOTAL	83	91	84	121		
My child has improved in mathematics:					3	7.76 a
-- yes	96 (94.1%)	95 (90.5%)	90 (88.2%)	117 (83.0%)		
-- no	6 (5.9%)	10 (9.5%)	12 (11.8%)	24 (17.0%)		
TOTAL	102	105	102	141		
My child's class took enough field trips					3	9.95*
-- yes	52 (52.5%)	59 (60.8%)	41 (43.2%)	89 (62.2%)		
-- no	47 (47.5%)	38 (39.2%)	54 (56.8%)	54 (37.8%)		
TOTAL	99	97	95	143		

* $p < .05$

** $p < .001$

a) not significant, $p = .052$

Discussion and Conclusions

This study collected substantial evidence which showed that varying class size in Grades 4 and 5 resulted in few changes in classroom functioning. Those effects which were identified as statistically significant seemed to be modest and the differences between class sizes were not dramatic.

The present study used various kinds of instruments to explore the effects of class size on a wide range of variables. To provide an overview of the findings it is easier to group the variables and present them as if several separate studies were done at the same time, using the same group of teachers and students and using the same research design.

The first set of variables examined teachers' expectations before the study and their perceptions and opinions after each year of the study.

Secondly, a great deal of data was collected about the activities of teachers and pupils in the various sized classrooms by using trained observers. From these data, it was, for instance, possible to examine how teachers and pupils spent their time and see what changes there were in teacher-pupil interaction. In addition it was possible to assess pupil participation, use of educational resources, method of instruction, subject emphasis, physical conditions and classroom atmosphere. These variables were assessed using an instrument developed as part of this project. Another observational scale called Indicators of Quality was also used.

Students' opinions and attitudes were collected each year and constitute a third area of the study. The instruments developed for this element of the study focused on physical conditions in the classroom, interpersonal relationships and general satisfaction with school as perceived by the students.

Student performance, a fourth aspect of the study, was assessed in two different ways. A standardized achievement test was used to measure Reading, Vocabulary, Mathematics-Concepts and Mathematics-Problem Solving. Samples of work done by the students were assessed to examine progress in Art and Composition. A measure of Self-Concept was also administered.

Finally, as a fifth element of the study, data were collected from parents by questionnaire to ascertain their attitudes and opinions.

Of all the different variables examined in this study, the ones that were most affected by class size were the teachers' expectations before the study and their attitudes and opinions after experiencing another class size. The teachers expected the two smaller classes (16 and 23) to be different from the larger classes (30 and 37), especially in the amount of individualization possible. Their experiences with the smaller classes as they perceived it supported this expectation. Even though the observational data do not seem to give these findings much support, they must not be ignored. It does make a positive difference to the teacher to work in a smaller class.

The investigation of observational variables assessing teacher-pupil interaction showed that few of them were affected by class size. In the smaller classes, that is sizes 16 or 23, the

teachers were less likely to speak to students about their behaviour, including all neutral, positive and negative comments. The neutral category was the only one to show a significant effect due to class size; the frequency was less in the smaller class. The differences in the amount of time spent talking about their pupils' behaviour were not large, though statistically significant. For example, teachers in classes of 30 or 37 spent 5 per cent of their time talking about pupils' behaviour, compared to 3.5 per cent in classes of 16 and 23. The differences observed in the frequency of teacher supervision were differences in a variable whose reliability could not be established; as well there was no further support for the finding as no significant differences existed in a very similar measure, the supervision variable, which was assessed differently and which was demonstrably reliable.

Class size did not affect the amount of time teachers spent talking about course content or classroom routines; it did not affect the amount of time they spent lecturing; it did not affect their choice of audience for what they said -- that is, teachers, when they changed sizes, did not alter the proportion of their time spent working with the class, with groups or with individuals. The differences in the frequency with which individual pupils were addressed by teachers were large, but as noted, there were no corresponding differences in the amount of time spent by the teacher talking to individual pupils. It seems, then, that pupils in classes of 16 and 23 had more individual interactions with their teachers simply because the amount of time spent talking to pupils individually was being distributed among fewer pupils. These findings demonstrate

support for some of the teachers' perceptions, but they are not the results of important changes in classroom management. Rather they are changes which would be expected if class sizes were reduced and teachers did not change their style of teaching.

Pupil participation was another area in which most variables were unaffected by class size. The observed differences in the extent of verbal participation were quite small, with students participating verbally 7.1 per cent of the time in smaller classes and 6 per cent of the time in larger classes. In general, pupils were participating 73.3 per cent of the time in the smaller classes and 72.7 per cent of the time in the larger classes -- a small and non-significant difference. Changing class size did not produce major changes in pupils' participation.

The only difference in the use of educational resources was in a variable whose reliability could not be established; the difference, furthermore, is not easy to explain as a result of class size. There were no significant differences in the quantities of educational resources seen in the different sizes. More interesting is the failure to confirm the teachers' opinions about the quantity of aids available in the different class sizes. Teachers tended to feel that the quantity of aids available in the smaller classes was superior, but the observational data report no differences in the numbers of aids available. The contradiction here may be only apparent, as the observational system counted only those aids which were displayed, rather than all aids which might have been available to the class. Another explanation is that the same number of resources may seem to be superior when there are fewer students to use them.

The only difference in method of instruction was found in the number of observation visits a year during which lecturing was used as a method of instruction. This particular variable, however, was not demonstrably reliable. A different assessment of the lecturing variable mentioned above, using a measure with demonstrated reliability, did not show any significant differences related to class size.

Reading was reported as having been taught during more observation visits in classes of 16 than in classes of other sizes. However, the actual length of time spent on reading, and other subjects, could not be assessed, and achievement in reading was not affected by class size.

Smaller classes were more likely than larger classes to be described as quiet, and less likely to be described as having "lots of noise." This finding seems logical as a small number of people would be expected to make less noise than a large number of people. However no data were collected which might indicate that noise levels were disruptive for any class sizes and there are no criteria for suitable classroom noise levels. The other physical characteristic of the classroom which was investigated -- furniture arrangement -- was not significantly related to class size. The failure to observe a difference again suggests that changes in class size did not produce important changes in classroom management, at least among most of the teachers. The classrooms with a smaller number of students had much more space for each student and the students' furniture. Because the study used existing classrooms, the average area of floor space available per student became greater as the number of students per teacher decreased.

Classroom atmosphere, as measured by an observational scale was not affected by class size. Indicators of Quality scores, also observational data, were affected by class size, but the effects that were recorded were weak compared to those reported in Olson's (1971) non-experimental study. It may be that the differences between class sizes observed by Olson were the result of several different variables, such as type of community, type of teacher and type of student, having been confounded with class size. In spite of the title, there is no direct evidence that this instrument measures what would be viewed as "quality education."

Considering the preceding, it appears that, in general, teachers experience a big difference between the smaller, 16 or 23, and the larger, 30 or 37, class sizes. This is perceived within the teacher's own personal frame of reference. They report making many changes in physical layout, evaluation, class management and supervision as a result of the differences (they were not instructed or advised as to how they should adjust for the different sizes). Undoubtedly, to an individual, a change personally initiated, for whatever reason, is seen to be important and significant. Viewed externally these individual changes are either so small as to be undetectable, or else they are observable but still are relatively small when seen in the context of the wide variety of other teachers' actions and activities.

Besides various measures of student performance, measures of student attitudes and opinions were also collected. Although initially it seemed that there were some class size effects, these apparent differences disappeared after introducing statistical controls to remove the variability due to teacher and year. Student

opinions on subject emphasis were also collected. The subjects listed were Art, Music, Gym, Reading, Writing and Mathematics. The students were asked each year to indicate whether the subject received too much, too little or just the right amount of time. Art showed an effect related to class size, such that the students in the small classes were less likely to say that Art received "too little" time than those in the larger classes. In general, for all sizes, there were more students who reported "too little" time for Art, Gym and Music than for the other three subjects. In addition, in all four class sizes there were more students who tended to feel that "too much" time was spent on Mathematics than there were who felt that way about Reading and Writing.

Standardized tests that were used to measure academic achievement showed a statistically significant class size effect in only one of the four areas assessed, Mathematics-Concepts. There were no significant differences found for the measures of Reading, Vocabulary or Mathematics-Problem Solving. Measures of the students' development in Art and Composition based on samples of the students' work showed no differences between class sizes. An argument against standardized tests was that performance in a creative endeavour such as Art or Composition would be more sensitive to class size effects than performance on a multiple choice test. This was not supported. Students showed no sign of development on the art measure over the two years.

These findings raise at least one possibility, namely that Mathematics is more sensitive to teacher availability than some of the other subjects. Mathematics, in comparison to Reading, is a

subject that likely gets a different kind of practice and amount of application outside the elementary classroom. Subjects such as Science and Social Studies were not assessed and so are not being considered. Art and Composition are interesting because the question raised by one observer after many observations was, "Are these taught or do the students just get an opportunity to exercise their present skills, receiving a useful comment from the teacher if they're lucky?"

There were no significant class size effects which could be identified in the measure of student's self-concept. This particular measure dealt with academic self-concept.

The parents' opinions and attitudes, collected at the conclusion of the study, provided another perspective on the study. The parents were considerably more accurate in judging the present size of their child's class the second year (71 per cent correct) than the size the first year (55.6 per cent). Parents preferred the smaller classes and tended to make positive comments about size 16 and 23 and negative comments about sizes 30 and 37. It is remarkable that 29.1 per cent of the parents did not think that class size affected the cost of education. It is possible to interpret the responses to these questionnaires as showing that parents have their educational awareness focused on the present, without necessarily considering the effect of changes beyond the way in which they might affect their child now. It must not be forgotten that smaller class sizes require more teachers, more classrooms and more money.

Considering the data carefully, it is easy to conclude that reducing class size is perceived as beneficial by teachers and

parents. There will be a few other effects that will likely make the classroom a slightly better place for the students and their arithmetic computation might show a just noticeable improvement, depending partly on how small the class becomes. The attitudes toward school, however, are unlikely to change, and the teachers' instructional style will not likely change either.

It is essential to note that these findings cannot be generalized to all levels and kinds of education. This study was based on students in junior grades and the results cannot be applied to primary grades, to kindergartens or to nursery schools. These results cannot be applied to secondary school or university either. In addition at the higher levels of education, classes usually operate on rotary timetables and this study cannot be generalized to such a setting.

The study avoided mixing the concept of class size with the concept of open plan schools and team teaching. In those situations, the traditional meaning of class size does not necessarily apply. In addition, there were no split grades in this study.

An important characteristic of this study which limits generalization of the results is that teachers were given no training in the use of appropriate methods to accomplish specified educational objectives with different class sizes. Training was not provided because it is unlikely, if class size were to be changed by legislation, that a well-defined in-service training program would be included in that legislation.

Essentially this study changed class size and nothing else. The researchers assumed that the influences of teachers and students were more important than class size and designed the study accordingly. Before further research is done it would be helpful to

have clear definitions of educational objectives and a set of strategies for attempting to reach them, differentiated by class sizes. Only then could appropriate training be provided and evaluated. At a special conference in 1975 various facets of the issues related to "class size" were presented (Ryan & Greenfield, 1976). These provide different kinds of perspectives than discussed here. It is the authors' hope that they will be able to use the data collected in this study to go beyond the class size question and explore other relationships such as those among teaching style, student attitudes and student performance.

References

Anderson, G. J. Effects of classroom social climate on individual learning. American Educational Research Journal, 1970, 7, pp. 135-152.

Anderson, G. J., & Walberg, H. J. Class size and the social environment of learning: a replication. Alberta Journal of Educational Research, 1972, 28(4), pp. 277-286.

Bales, R. Interaction process analysis. Reading, Mass.: Addison-Wesley, 1950.

Balow, I. H. A longitudinal evaluation of reading achievement in small classes. Washington, D.C.: U.S. Department of Health, Education and Welfare, Office of Education, 1967. (ERIC Document Reproduction Service No. ED 011 813, RE 000 121.)

Beere, C. A. Development of a group instrument to measure young children's attitudes toward school. Psychology in the Schools, July, 1973, pp. 308-315.

Blake, H. E. Class size: a summary of selected studies in elementary and secondary public schools. Unpublished doctoral dissertation, Teachers College, Columbia University, 1954.

Blishen, B. R. A socio-economic index for occupations in Canada. Canadian Review of Sociology and Anthropology, 1967, 4(1), pp. 41-53.

British Columbia School Trustees Association. Class size and pupil-teacher ratio: an annotated bibliography. Vancouver: British Columbia School Trustees Association, 1974.

British Columbia Teachers' Federation. Olson provides ammunition for smaller classes. British Columbia Teachers' Federation Newsletter, 1973, 13(1), p. 1.

Cannon, G. M. Kindergarten class size -- a study. Childhood Education, September, 1966, pp. 9-11.

Carr, W. G. New angle of attack needed on class size research. Nations Schools, 1932, 10, pp. 27-30.

Cogan, M. The relation of the behaviour of teachers to the productive behaviour of their pupils. Journal of Experimental Education, 1958, 27, pp. 89-124.

Cohen, J. Statistical power analysis for the behavioural sciences. New York: Academic Press, 1969.

Coleman, J. S. Equality of educational opportunity. Washington, D.C.: U.S. Department of Health, Education and Welfare, Office of Education, 1966.

Cronbach, L. J. Essentials of psychological testing (2nd ed.). New York: Harper & Row, 1970.

Danowski, C. E. Individualization of instruction: a functional definition. IAR Research Bulletin, 1965, 5(2), pp. 1-8.

Deosaran, R. A., Wright, E. N., & Kane, T. The 1975 every student survey - student's background and its relationship to program placement. Toronto: The Board of Education for the City of Toronto, Research Department, 1976 (#138).

Edeburn, C. E. & Landry, R. G. Teacher self concept and student self concept in grades three, four and five. Journal of Educational Research, 1976, 69, pp. 372-375.

Flanders, N. Analyzing teaching behaviour. Reading, Mass.: Addison-Wesley, 1970.

Fox, R., Luszki, M., & Schmuck, R. Diagnosing classroom learning environments. Chicago: Science Research Associates, 1966.

Gaitskell, C., & Hurwitz, A. Children and their art: Methods for the elementary school. (2nd ed.) New York: Harcourt, Brace & World, 1970.

Gajewsky, S., & Shore, B. M. Class size - review of the literature and selected annotated bibliography. Reports in Education (Monograph Series). Montreal: McGill University, Departments of Educational Administration and Educational Psychology and Sociology, Faculty of Education, 1973.

Hall, D. T. The effect of teacher-student congruence upon student learning. New York: John Wiley, 1966.

Harris, B. M. A research study of the effects of demonstration teaching upon experienced and inexperienced teachers. Unpublished doctoral dissertation, University of Texas at Austin, 1966.

Herbert, J. A system for analyzing lessons. New York: Teachers College Press, 1967.

King, E. M. (ed.). Canadian tests of basic skills: Teacher's manual. Toronto: Thomas Nelson & Sons, 1968.

Kirby, D. M., & Gardner, R. C. Ethnic stereotypes: norms on 208 words typically used in their assessment. Canadian Journal of Psychology, 1972, 26(2), pp. 140-154.

Lorber, N. M. Measuring the character of children's peer relations using the Ohio Social Acceptance Scale. California Journal of Educational Research, 1969, 24(2), pp.71-77.

Lindbloom, D. H. Class size as it affects instructional procedures and educational outcomes. Minneapolis: Educational Research and Development Council of the Twin Cities Metropolitan Area, 1970. (ERIC Document Reproduction Service No. ED 059 532.)

Medley, D., & Mitzel, H. Measuring classroom behaviour by systematic observaton. In N. L. Gage (Ed.) Handbook of research on teaching. Chicago: Rand McNally, 1963.

Mosteller, F., & Tukey, J. Data analysis, including statistics. In G. Lindzey & E. Aronson (Eds.) The handbook of social psychology, (Second Edition, Volume 2), Reading, Mass.: Addision-Wesley, 1968, pp. 80-203.

Oliva, F. D., Maguire, T. O., & Pacey, R. J. IPI project report. Unpublished manuscript, Human Resources Research Council, Edmonton, Alberta, June, 1971.

Olson, M. Classroom variables that predict school system quality. I A R Research Bulletin, 1970, 11(1), pp. 1-11.

Olson, M. Identifying quality in school classrooms: some problems and some answers. Exchange, 1971 (Jan.), 29(5), pp. 1-11.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. The measurement of meaning. Urbana, Ill.: University of Illinois Press, 1961.

Pennington, S. Children's drawings: a selected review of literature. Toronto: The Board of Education for the City of Toronto, Research Department, 1967 (#43).

Purkey, W. W. Self concept and school achievement. New Jersey: Prentice Hall, 1970.

Report on Education Research. Teachers speak out on their problems. Report on Education Research, October 11, 1972, p. 11.

Ryan, D. W., & Greenfield, T. B. The class size question: development of research studies related to the effects of class size, pupil-adult, and pupil-teacher ratios. Toronto: Ministry of Education, 1975.

Ryan, D. W., & Greenfield, T. B. Clarifying the class size question: evaluation and synthesis of studies related to the effects of class size, pupil-adult, and pupil-teacher ratios. Toronto: Ministry of Education, 1976.

Shapiro, E. Study of children through observation of classroom behaviour. In A. Bellack (Ed.) Theory and research in teaching. New York: Teachers College, Columbia University, 1963.

Shapson, S. Optimum class size? A review of the literature. Toronto: The Board of Education for the City of Toronto, Research Department, 1972 (#114).

Shapson, S. M., Virgin, A. E., & Crawford, P. Development of an instrument to measure self concept in schools. Toronto: Board of Education for the Borough of North York, Educational Research Services, October, 1971.

Sitkei, E. G. The effects of class size: A review of the research. Washington, D.C.: U.S. Department of Health, Education and Welfare, 1968.

Steele, J. M., House, E. R., & Kerins, T. An instrument for assessing instructional climate through low inference student judgments. American Educational Research Journal, 1971, 8, pp. 447-466.

Stephens, L., & Berryman, C. The effects of class size and instructional method on attitudes toward the teaching profession. Contemporary Education, 1968, 40(2), pp. 76-78.

Stones, E. Students' attitudes to the size of teaching groups. Educational Review, February, 1969, pp. 98-108.

Vincent, W. Further clarification of the class size question. IAR Research Bulletin, 1968, 9(1), pp. 1-3.

Vincent, W. Class size. In R. L. Ebel (Ed.), Encyclopedia of Educational Research. Toronto: Macmillan, 1969, pp. 142-146.

Vincent, W., & Olson, M. Measurement of school quality and its determiners. New York: Teachers College, Columbia University, 1972.

Wahlstrom, M. W. The class activities questionnaire. Toronto: Ontario Institute for Studies in Education, 1971.

Walberg, H. J., & Ahlgren, A. Predictors of the social environment of learning. American Educational Research Journal, 1970, 7(2), pp. 153-167.

Wright, E. N., & Reich, C. Language: a study of fundamental skills. Toronto: Board of Education for the City of Toronto, Research Department, 1972, (#108).

Yamamoto, K., Thomas, E. C., & Karns, E. A. School related attitudes in middle-school age students. American Educational Research Journal, 1969, 6, pp. 191-206.